Universidad Nacional Mayor de San Marcos
Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado
Facultad de Ciencias Matemáticas
Unidad de Posgrado

Distribución asintótica de los estimadores MCO en una regresión lineal con variables explicativas que siguen procesos estocásticos strong-mixing y tendencia

TESIS

Para optar el Grado Académico de Magíster en Estadística Matemática

AUTOR

Juvert Alexi HUARANGA NARVAJO

ASESOR

Mg. Antonio BRAVO QUIROZ

Lima, Perú

2018

Referencia bibliográfica

Huaranga, J. (2018). *Distribución asintótica de los estimadores MCO en una regresión lineal con variables explicativas que siguen procesos estocásticos strong-mixing y tendencia*. [Tesis de maestría, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Matemáticas, Unidad de Posgrado]. Repositorio institucional Cybertesis UNMSM.

ACTA DE SUSTENTACIÓN DE TESIS DE GRADO ACADÉMICO DE MAGÍSTER

Siendo las, 17.15 horas del día miércoles treinta y uno de octubre del dos mil dieciocho, en el Aula 105 de la Facultad de Ciencias Matemáticas, el Jurado Evaluador de Tesis, Presidido por el Mg. Wilfredo Eugenio Domínguez Cirilo e integrado por los siguientes miembros, Mg. Emma Norma Cambillo Moyano (Jurado Evaluador), Mg. Violeta Alicia Nolberto Sifuentes (Jurado Evaluador), Mg. Rosa Fátima Medina Merino (Jurado Evaluador) y el Mg. Antonio Bravo Quiroz como Miembro Asesor, se reunieron para la sustentación de la tesis titulada: «DISTRIBUCIÓN ASINTÓTICA DE LOS ESTIMADORES MCO EN UNA REGRESIÓN LINEAL CON VARIABLES EXPLICATIVAS QUE SIGUEN PROCESOS ESTOCÁSTICOS STRONG-MIXING Y TENDENCIA» presentada por el Bachiller Juvert Alexi Huaranga Narvajo, para optar el Grado Académico de Magíster en Estadística Matemática.

Luego de la exposición del graduando, los Miembros del Jurado hicieron las preguntas correspondientes, así como las observaciones e inquietudes acerca del trabajo de tesis, a las cuales el Bachiller Juvert Alexi Huaranga Narvajo respondió con acierto y solvencia, demostrando pleno conocimiento del tema.

A continuación se realizó la calificación correspondiente, según tabla adjunta, resultando el Bachiller Juvert Alexi Huaranga Narvajo aprobado con el calificativo de BUENO ....QUINCE................

Habiendo sido aprobada la sustentación de la Tesis, el Jurado Evaluador recomienda para que el Consejo de Facultad apruebe el otorgamiento del grado académico de **Magíster en Estadística Matemática** al Bachiller **Juvert Alexi Huaranga Narvajo**.

Siendo las 18.10 horas, se levantó la sesión, firmando para constancia la presente Acta.

Mg. Emma Norma Cambillo Moyano
Miembro

Mg. Wilfredo Eugenio Domínguez Cirilo
Presidente

Mg. Violeta Alicia Nolberto Sifuentes
Miembro

Mg. Rosa Fátima Medina Merino
Miembro

Mg. Antonio Bravo Quiroz
Miembro Asesor

Distribución asintótica de los estimadores MCO en una regresión lineal con variables explicativas que siguen procesos estocásticos strong-mixing y tendencia

Juvert Alexi Huaranga Narvajo

tesis presentada a consideración del cuerpo docente de la escuela Profesional de Estadística de la Facultad de Ciencias Matemáticas de la Universidad Nacional Mayor de San Marcos para obtener el grado de Magister en Estadística Matemática.

Aprobado por:

Mg. Wilfredo Eugenio Dominguez Cirilo
Presidente

Mg. Emma Norma Cambillo Moyano
Miembro

Mg. Violeta Alicia Nolberto Sifuentes
Miembro

Mg. Rosa Fátima Medina Merino
Miembro

Mg. Antonio Bravo Quiroz
Miembro Asesor

FICHA CATALOGRÁFICA

HUARANGA NARVAJO, JUVERT ALEXI

Distribución asintótica de los estimadores MCO en una regresión lineal con variables explicativas que siguen procesos estocásticos strong-mixing y tendencia, (Lima) 2018.

VIII, 90p, 29.7cm(UNMSM, Magister, Estadística Matemática, 2018)

Tesis de Maestría, Universidad Nacional Mayor de San Marcos. Facultad de Ciencias Matemáticas , Escuela de estadística. Unidad de Posgrado, UNMSM/FCM

Dedicatoria:

Dedico esta tesis a Dios por darme la vida cada día, a mis profesores de maestría por compartir sus invaluables conocimientos y a esta prestigiosa casa de estudios por la oportunidad brindada de seguir mejorando.

Presentación

En el estudio de fenómenos sociales y naturales es de vital importancia hallar las relaciones causales entre las variables y cuantificar el impacto que cada parámetro tiene dentro de un modelo o sistema debido a las implicaciones teóricas, en el desarrollo de políticas y en la elaboración de pronósticos. Por consiguiente los procedimientos de inferencia, que van a determinar si existen o no relaciones causales y en que magnitud, cobran especial importancia a la hora estudiar un fenómeno.

En el contexto de las series temporales, y al igual que con otros tipos de datos, los procedimientos para determinar la significancia o no de una determinada variable en un modelo o sistema hacen uso de test estadísticos y sus funciones de distribución. Debido a las propiedades estocásticas particulares de diversos tipos de series temporales es común encontrar relaciones causales en donde no las hay como consecuencia de ignorar la verdadera distribución que siguen los estimadores y el test estadístico bajo determinado tipo de proceso estocástico y en su lugar usar distribuciones estándar de los test estadísticos para hallar la significancia de cada variable.

En consecuencia el presente trabajo tiene por objetivo obtener la distribución asintótica de los estimadores de mínimos cuadrados ordinarios (MCO) en regresiones de series temporales con variables explicativas que siguen procesos estocásticos strong-mixing y tendencia. De modo que puedan ser usados para hallar relaciones causales válidas entre las variables y poder así dar respuestas válidas a los problemas de interés que tengan los investigadores en los diversos campos de investigación que hacen uso intensivo de series temporales.

Introducción

Debido a la existencia de fenómenos que siguen procesos no estacionarios que aparecen en diversas disciplinas como: astronomía, química, economía, finanzas, biología y entre otros, la teoría estadística convencional no es la adecuada para realizar inferencias, sobre los diversos fenómenos que se tratan de explicar, dado que en el contexto de series temporales supone la existencia de procesos estacionarios que sean ergódicos para poder aplicar las herramientas de inferencia convencionales ya sea variables univariadas o multivariadas, en modelos estructurales, regresiones, sistemas dinámicos, etc.

En el contexto de las regresiones con series temporales el supuesto de ergodicidad permite que la distribución los estimadores de mínimos cuadrados ordinarios (MCO, en adelante) sigan una distribución asintótica normal (Durlauf y Phillips (1988)) y tanto las pruebas: t-student y F-Snedecor sigan distribuciones asintóticas t y F respectivamente. Sin embargo es común que que las series encontradas sigan procesos no estacionarios, en tal situación la aplicación del método MCO en regresiones da como resultado que los estimadores MCO sigan una distribución aleatoria mientras que los las pruebas t-student y F-Snedecor siguan distribuciones sesgadas. En tal contexto los procedimientos de inferencia van a mostrar resultados erróneos estableciendo relaciones causales en donde no las hay.

En consecuencia surgió la necesidad de hacer nuevos desarrollos en la teoría de la probabilidad que puedan proporcionar las herramientas analíticas adecuadas para obtener las distribuciones asintóticas de los estimadores MCO en regresiones con variables que siguen procesos no estacionarios. La obtención de tales distribuciones está basada en distribuciones de Wiener, el teorema del límite central funcional (TLCF) y varios teoremas del análisis real y funcional.

No obstante los desarrollos en teoría de la probabilidad dados en décadas recientes, no es posible hallar de manera general la distribución de los estimadores MCO en regresiones con variables no estacionarias. En cambio, se han hallado distribuciones asintóticas en regresiones con variables que siguen procesos específicos AR(1), AR(2), procesos integrados de orden (1), orden (2), estacionarios en diferencia de orden(1), entre otros. Esto se debe a que el comportamiento de los procesos estocásticos en mención varía, enormemente en algunos casos, cuando varía el valor de los parámetros. Por lo que existe espacio para desarrollos posteriores en la obtención de distribuciones asintóticas usando diferentes tipos de procesos. Tal es el caso del presente trabajo que tiene como objetivo obtener la distribución asintótica del estimador OLS cuando los regresores siguen procesos integrados y estacionarios en tendencia.

Finalmente, el presente trabajo se dividide en 5 capítulos. El primer capítulo aborda el tema de investigación de la presente tesis: situación problemática, justificación y objetivos así como la revisión de la literatura. En el segundo capítulo se realiza una exposición de las definiciones y resultados fundamentales en teoría de la probabilidad e inferencia estadística asi como de análisis real y análisis funcional necesarios para derivar los resultados estadísticos de la presente tesis. El capítulo tres describe el modelo de regresión lineal múltiple; se hallan los estimadores del mismo bajo el método de los mínimos cuadrados ordinarios, se presenta sus propiedades estadísticas y posterior uso en inferencia. En el capítulo cuatro se deriva de manera rigurosa los resultados de la presente investigación, esto es, la distribución asintótica de los estimadores MCO en una regresión lineal con variables explicativas que siguen procesos estocásticos strong-mixing y tendencia. Y en el último capítulo se presentan las conclusiones y recomendaciones.

Índice general

Capítulo 1

La presente investigación

1.1. Situación Problemática

El estudio y análisis de las series temporales tiene aplicaciones en biología, economía, ingeniería, astronomía y otros campos por lo que se ha convertido en un área de constante crecimiento dentro de la estadística matemática. Dado que la naturaleza de las series temporales es dinámica, uno de los objetivos básicos en el trabajo aplicado es identificar estas relaciones dinámicas entre las series temporales en caso exista.

Existen diversos métodos para la obtención de relaciones dinámicas entre las series temporales, siendo los más usados: modelos espacio-estado, vectores autoregresivos, funciones de transferencia y regresiones. En comparación a los demás métodos el uso de regresiones en el análisis de series temporales es de fácil interpretación y programación en paquetes estadísticos, en contraparte, las distribuciones de los estimadores y test estadísticos en una regresión de series temporales generalmente son asintóticas y no son estándar debido a que las series temporales son procesos no dependientes.

La necesidad de obtener la distribución de los estimadores en regresiones de series temporales dio desarrollo a la teoría estadística necesaria para su obtención, de modo que los estimadores y test estadísticos en cada regresión con diferentes tipos de variables o combinación de variables tienen su propia distribución. Esta teoría está basada en regresiones estimadas mediante el método de mínimos cuadrados ordinarios (MCO), tal método consiste en minimizar la sumatoria de los cuadrados del error con respecto a los parámetros para obtener los estimadores (Seber y Lee, 2003).

1.2. Justificación de la investigación

La investigación se justifica porque al conocer la verdadera distribución de los estimadores MCO en una regresión lineal con variables explicativas que siguen procesos estocásticos strong-mixing y tendencia permite implementar procedimientos inferenciales más generales en el marco del modelo de regresión ya que se pueden al incluir variables dependientes con propiedades estadísticas más débiles que las requeridas en otros modelos de regresión y al mismo tiempo se puede incluir una tendencia en el modelo, y en consecuencia evitar las relaciones causales espurias en regresiones de series temporales no estacionarias, esto es, sostener que existe una relación causal entre las variables cuando en realidad no la hay. Tal procedimiento es de gran importancia en trabajos aplicados.

1.3. Objetivos de la investigación

1.3.1. Objetivo general

- Derivar analíticamente la distribución de los estimadores MCO en una regresión con tendencia y variables explicativas que siguen procesos estocásticos dependientes strong-mixing.

1.4. Revisión de la literatura

El estudio de las propiedades asintóticas de los estimadores MCO en regresiones con series temporales es de larga data. En el plano univariado, se tienen los trabajos de Park y Phillips (1988) que desarrollan la teoría asintótica en regresiones multivariantes con procesos integrados de diferente orden, tales regresiones pueden tener interceptos, derivas y tendencias temporales. En Durlauf y Phillips (1988) se analizan las propiedades asintóticas de los estimadores MCO cuando se realiza la extracción errónea de la tendencia determinística en procesos que son integrados. Mientras que en Haldrup (1994) se deriva las propiedades de los estimadores MCO en regresiones que cointegran con variables que siguen procesos integrados de orden uno y dos. Además en Hasseler (2000) son derivadas las propiedades asintóticas de los estimadores MCO en regresiones con procesos que siguen tendencias lineales. Y en Phillips (1987) se presenta las distribuciones tanto de los estimadores MCO como de los test estadísticos en regresiones cuyas variables siguen un proceso auto-regresivo. Por otro lado Kramer y Marmol (2002) se derivan las propiedades asintóticas de los estimadores MCO y mínimos cuadrados generalizados (GLS, en adelante) en regresiones lineales donde los errores siguen un proceso autoregresivo (AR) de orden p.

En un contexto multivariado, se tiene a Phillips y Durlauf (1986) se desarrolla la teoría asintótica para variables integradas de orden uno en regresiones multivariadas, vectores auto-regresivos y modelos de corrección del error; y bajo diferentes tipos de errores, entre ellos: errores con dependencia débil y heterogeneidad, en el proceso generador de datos (PGD, en adelante). Mientras que en Stock, Sims y Watson (1990) que desarrollan las propiedades asintóticas y de las pruebas de hipótesis de los estimadores MCO cuando algunas de las variables siguen procesos integrados de diferente orden. Por otro lado en Park y Phillips (1989) se analiza las propiedades asintóticas de variables que siguen procesos integrados de orden uno en regresiones multivariadas así como las distribuciones de los estadísticos en diversas pruebas de hipótesis.

Así también, un área de investigación intrínseca al de las distribuciones asintóticas de estimadores MCO en regresiones es el de la causalidad espuria, en el cual también se generan varios resultados asintóticos de interés. Así se tienen los trabajos de Hassler (1996) donde deriva las distribuciones asintóticas de los estimadores MCO en una regresión donde el regresor es integrado de orden uno e independiente del regresando que sigue un proceso estacionario, además logra establecer la causalidad espuria entre las variables. Entorf (1997) halla la distribución asintótica del estimador MCO en una regresión de dos variables independientes que siguen procesos de caminata aleatoria con deriva. Mientras que en Stewart (2006) se hallan las distribuciones límite de los estimadores MCO y de los test estadísticos convencionales en regresiones univariadas con variables estacionarias e integradas. Por otro lado en Noriega y Ventosa-Santaularia (2007) se descubre causalidad espurea en una regresión con variables que siguen procesos con tendencia lineal y desplazamiento de tendencia, así mismo halla las distribuciones asintóticas de los estimadores.

De modo que la literatura es extensa y variada en el área en cuestión, dejando mucho espacio para desarrollos posteriores.

Capítulo 2

Bases teóricas y Marco teórico

2.1. Fundamentos de analisis real y funcional

Definición 2.1.1. σ-álgebra. Una colección de subconjuntos $\mathcal{A}$ del conjunto X se llama σ-**álgebra**[1], si cumple las siguientes condiciones:

1. $X \in \mathcal{A}$
2. $A \in \mathcal{A} \Longrightarrow A^c \in \mathcal{A}$
3. $A, B \in \mathcal{A} \Longrightarrow A \cup B \in \mathcal{A}$
4. $(A_n : n \in \mathbb{N}) \subset \mathcal{A} \Longrightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$

Definición 2.1.2. Espacio medible. Sea $\mathcal{A}$ un $\sigma - algebra$ de los subconjuntos del conjunto X. El par $\big(X, \mathcal{A}\big)$ es llamado un espacio medible[2].

Definición 2.1.3. Medida. Sea $\mathcal{A}$ un $\sigma - algebra$ de los subconjuntos del conjunto X. Una funcion de conjuntos μ definida en $\mathcal{A}$ se llama medida[3] si satisface las siguientes condiciones:

[1] Una completa explicación del concepto de σ-álgebra y sus propiedades se encuentra en Yeh (2006), cap-1.

[2] Ibid.

[3] Ibid.

1. $\mu\big(E\big) \in \big[0, \infty\big\rangle$ *para todo* $E \in \mathcal{A}$
2. $\mu\big(\emptyset\big) = 0$
3. $\big(E_n : n \in \mathbb{N}\big) \subset \mathcal{A}$, *disjunto* $\Longrightarrow \mu\Big(\bigcup_{n\in\mathbb{N}} E_n\Big) = \sum_{n\in\mathbb{N}} \mu\big(E_n\big)$

Definición 2.1.4. Espacio medido. Si μ es una medida en un $\sigma - algebra$ $\mathcal{A}$ de un conjunto de subconjuntos de X. El triple $\big(X, \mathcal{A}, \mu\big)$ recibe el nombre de espacio medido[4].

Definición 2.1.5. Función Medible. Sean $(X, \mathcal{A})$ y $(Y, \mathcal{B})$ dos espacios medibles, una función $f \colon X \longrightarrow Y$ se dice que es $(\mathcal{A}, \mathcal{B})-$medible[5] si $f^{-1}(\mathcal{B}) \subset \mathcal{A} \ \ \forall \ \ B \in \mathcal{B}$

Definición 2.1.6. Espacio Métrico. Un conjunto no vacio X con una función d : $X \times X \longrightarrow \mathbb{R}$ se le llama espacio métrico si la función tiene las siguientes propiedades[6]

1. $d\,(x, y) \geq 0 \; x, y \in X$
2. $d\,(x, y) = si\,y\,solo\,si\; x = y$
3. $d\,(x, y) = d\,(y, x) \;\; x, y \in X$
4. $d\,(x, y) \leq d\,(x, z) + d\,(z, y) \;\; x, y, z \in X$

La función d recibe el nombre de metrica en X o algunas veces función de distancia en X. De modo que cuando se tengan expresiones del tipo (X, d) nos estamos refiriendo a un espacio métrico en donde d es una métrica del conjunto X.

Definición 2.1.7. El espacio métrico de funciones continuas. Sea X el conjunto de todas las funciones continuas definidas en el intervalo $[a, b]$, un invervalo en $\mathbb{R}$, para cada $f, g \in X$, se define:

$$d\,(f, g) = \sup_{x\in[a,b]} \big|f\,(x) - g\,(x)\big|$$

[4]Ibid.
[5]Vease Schilling (2005).
[6]Shirali (2006) p-27.

Se puede demostrar que d es una medida en X y que cumple todas las propiedades enumeradas en la definición anterior por lo que la expresión $\mathcal{C}\,[a,b]$ denota al espacio métrico de todas las funciones continuas[7].

Definición 2.1.8. El espacio métrico $\mathcal{D}\,[0,1]$. Sea X el espacio de todas las funciones continuas desde la derecha y con límites izquierdos definidas tal que $f:[0,1]\longrightarrow\mathbb{R}$ en el intervalo $[0,1]$, un invervalo en $\mathbb{R}$, Para cada $x,y\in X$ se define la métrica[8]:

$$d\,(x,y) \coloneqq \text{ínf}\left\{\varepsilon>0:\exists\lambda\in\mathit{\Lambda}\mid\|\lambda\|<\varepsilon\, y\ \sup_{t}|x(t)-y\,(\lambda\,(t))\leq\varepsilon|\right\}$$

Donde: $\|\lambda\|\coloneqq\sup_{s\neq t}|\log\lambda\,((t)-\lambda\,(s))\,/\,(t-s)|$ y $\mathit{\Lambda}\coloneqq[0,1]\longrightarrow[0,1]$ y siendo λun mapeo estricto creciente continuo de $[0,1]$ a sí mismo.

Definición 2.1.9. Espacio L^p. Sea $\big(X,\mathcal{M},\mu\big)$ un espacio medido. Se denota por $L^p\big(X,\mathcal{A},\mu\big)$ o simplemente $L^p\big(\mu\big)$ al conjunto de funciones medible definidas en X tal que[9]:

$$\int_X |f|^p\,d\mu<\infty$$

De manera equivalente, para $p\in(0,\infty)$, se define la $p-$norma como:

$$\|f\|_p=\left(\int_X |f|^p\,d\mu\right)^{1/p}$$

El espacio L^p es la coleción de todas las funciones medibles f en X con $\|f\|_p<\infty$[10].

[7]Ibid, p-32
[8]Véase Mishura (2008), p-80.
[9]Bruckner y Thomson (1997).
[10]Yeh (2006).

2.2. Fundamentos de probabilidad e inferencia

2.2.1. Conceptos básicos de probabilidad

Definición 2.2.1. Elemento aleatorio. Sea $(\Omega, \mathcal{F})$ un espacio medible y (S,s) un espacio métrico medible, una función medible X es un elemento aleatorio[11] si:

$$X : (\Omega, \mathcal{F}) \longrightarrow (S,s)$$

Definición 2.2.2. Función aleatoria. Sea $(\Omega, \mathcal{F})$ un espacio medible y $\mathcal{C}[a,b]$ el espacio métrico de funciones reales y continuas en el intervalo $[a,b]$, entonces la función medible X es una función aleatoria[12] si:

$$X : (\Omega, \mathcal{F}) \longrightarrow \mathcal{C}[a,b]$$

Definición 2.2.3. Espacio de probabilidad. Sea Ω el espacio muestral con un $\sigma-$algebra S de subconjuntos de Ω. Y sea P una medida de probabilidad en S con $P(\Omega) = 1$. Entonces el triple $(\Omega, \mathcal{F}, P)$ recibe el nombre de espacio de probabilidad[13].

Definición 2.2.4. Variable Aleatoria. Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. Una variable aleatoria[14] es una función medible $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$[15]. Y se denota como:

$$X : \mathit{\Omega} \longrightarrow \mathbb{R}$$

Esto significa que una variable aleatoria asigna a cada elemento del espacio muestral un valor del conjunto de los números reales.

Definición 2.2.5. Vector Aleatorio. Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad, un

[11] Véase Gut (2013) p-45.
[12] Ver Billinsley (1968) p-22.
[13] Ver Dudley (2002) para mayores detalles
[14] Mayores detalles en Borovkov (2013), cap-3
[15] Donde $\mathcal{B}$ es el σ-álgebra de todos los conjuntos de Borel.

vector aleatorio[16] n-dimensional $\mathbf{X}$ es una función medible desde el espacio muestral Ω hacia $\mathbb{R}^n$, esto es:

$$\mathbf{X} = \Omega \longrightarrow \mathbb{R}^n$$

Definición 2.2.6. Función de distribución. Para una variable aleatoria X en un espacio de probabilidad $\big(\Omega, \mathcal{F}, P\big)$, se define $F_x\big(X\big)$ como la probabilidad de que x no exceda X, esto es:

$$F_x\big(X\big) = P\Big(\big\{\omega : x\left(\omega\right) \leq X\big\}\Big) \; X \in \mathbb{R}$$

La funcion $F_x\big(X\big)$ se llama funcion de distribucion[17] de la variable aleatoria X.

Definición 2.2.7. Sucesión de variables aleatorias. El conjunto ordenador de variables aleatorias $X_1, X_2, \ldots X_n$ se llama sucesión de variables aleatorias y se denota como $\{X_n\}$.

Definición 2.2.8. Sucesión de variables aleatorias I.I.D. Sea$\{X_n\} = \big\{X_1, X_2, \ldots X_n\big\}$ una sucesión de variables aleatorias, se dice que una sucesión de variables aleatorias es independiente[18] si:

$$F_{X_1 X_2 \cdots X_n}(X_1 X_2 \cdots X_n) = F_{x_1}(X_1) \times F_{x_2}(X_2) \times \ldots \times F_{x_n}(X_n)$$

Y es identicamente distribuida si:

$$F_{x_1}(X_1) = F_{x_2}(X_2) = \ldots = F_{x_n}(X_n)$$

En consecuencia, en una sucesión de variables aleatorias I.I.D todas las variables aleatorias son independientes unas de otras y tienen la misma función de distribución[19] de modo que la función de distribución de una sucesión de variables aleatorias

[16] Mayores detalles en Roussas (2014), pag-8.
[17] Mayores detalles en Basu (2012), p-25.
[18] Véase Athreya y Lahiri (2006), p-208.
[19] De lo anterior se deduce que una sucesión de variables aleatorias $\{X_n\}$ es heterogenea si las

I.I.D se puede expresar como:

$$F_{X_1X_2\cdots X_n}(X_1X_2\cdots X_n) = \prod_{i=1}^{n} F_{x_i}(X_i) = \prod_{i=1}^{n} F_{x_1}(X_1)$$

Definición 2.2.9. Expectativa de una variable aleatoria. Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad, la expectativa de una variable aleatoria[20] es la integral de X con respecto a la medida P , es decir:

$$E[X] = \int XdP = \int_{\Omega} X(\omega)\, P(d\omega)$$

Proposición 2.2.1. *Una variable variable aleatoria X es integrable si y solo si $E[|X|] < \infty$[21].*

Definición 2.2.10. Momento y momento Absoluto. Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad, el $k-esimo$ momento absoluto[22] de una variable aleatoria X se define como:

$$E[|X|^k] = \int_{-\infty}^{\infty} |X|^k dF(x)$$

Y el $k - esimo$ momento de una variable aleatoria X se define como:

$$E[X^k] = \int_{-\infty}^{\infty} X^k dF(x)$$

Proposición 2.2.2. *Relación de entre el momento y momento absoluto de una*

variables aleatorias que la conforman siguen distintas distribuciones.

[20] Ver Billinsley (1995), p-273.

[21] Ibid.

[22] Ibid, p-274.

variable aleatoria. Sea X una variable aleatoria, entonces se cumple lo siguiente[23]*:*

$$(i)\ \ E\left[\left|X\right|\right] < \infty \Longleftrightarrow E\left[X\right] \qquad existe\, y\, es\, finita$$

$$(ii)\ \ E\left[\left|X\right|^{s}\right] < \infty \Longrightarrow E\left[\left|X\right|^{r}\right] < \infty \qquad para\, 0 < r \leq s$$

$$(iii)\ \ E\left[\left|X\right|^{s}\right] < \infty \Longrightarrow E\left[X^{k}\right] \qquad existe\, y\, es\, finita\,,\ 0 < k \leq s$$

Proposición 2.2.3. *Desigualdad de Jensen. Sea X una variable aleatoria, g una función convexa y que X y $g\left(X\right)$ son integrables, entonces*[24]*:*

$$g\left(EX\right) \leq Eg\left(X\right)$$

Definición 2.2.11. Espacio L^{p} para variables aleatorias[25]. Sea $\left(\Omega, \mathcal{F}, P\right)$ un espacio de probabilidad, el espacio $L^{p}\left(P\right)$ es la colección de todas las variables aleatorias $X : \Omega \longrightarrow \mathbb{R}$ que son integrables de orden P, esto es que: $E\left[\left|X\right|^{P}\right] < \infty$. De manera equivalente, para $p \in (0, \infty)$, se define:

$$\left\|X\right\|_{p} := \left[\left\{E\left|X\right|^{p}\right\}\right]^{1/p}$$

Una variable aleatoria X es P integrable si y solo si $\left\|X\right\|_{p} < \infty$.

Proposición 2.2.4. *Desigualdad de Holder. Sea $p^{-1} + q^{-1} = 1$. Si $E\left|X\right|^{p} < \infty$ y $E\left|Y\right|^{p} < \infty$, entonces*[26]*:*

$$\left|E\left[XY\right]\right| \leq E\left[\left|XY\right|\right] \leq \left\|X\right\|_{p} \times \left\|Y\right\|_{q} \tag{2.2.1}$$

Definición 2.2.12. Sucesión de variables aleatorias acotadas en $L^{p}\left(P\right)$. Una suce-

[23] Estas y otras propiedades se encuentran en Loeve(1978), p-157.
[24] Véase Gut (2013) p-132.
[25] Véase Khoshnevisan (2007) pag-39.
[26] Véase Gut (2013) p-129 para una demostración de este resultado.

sión de variables aleatorias $\{X_n\}$ es acotada en $L^p(P)$[27] si:

$$\sup_t E|X_t|^p < \infty$$

Definición 2.2.13. Acotamiento estocástico. Una sucesión de variables aleatorias $\{X_n\}$ es acotada estocásticamente[28] si, para cada $\varepsilon \in (0,1)$, existe un $M > 0$ finito tal que:

$$\inf_{n\geq 1} P\Big[\|X_n\| \leq M\Big] > 1 - \varepsilon$$

El acotamiento estocástico se denota por $O_p(1)$, de modo que $X_n = O_p(1)$ significa que la sucesión de variables aleatorias $\{X_n\}$ es acotada estocásticamente.

Proposición 2.2.5. *Si* $\sup_n E|X_n|^p < \infty$*, entonces* $X_n = O_p(1)$*. Esto es, si* $\{X_n\}$ *es acotada en* $L^p(P)$ *entonces es acotada estocásticamente*[29].

Definición 2.2.14. Sucesión de variables aleatorias uniformenente integrable. Una sucesión de variables aleatorias es uniformenete integrable[30] si:

$$\lim_{\alpha\longrightarrow\infty} \sup_n \int_{(|X_n|\geq\alpha)} |X_n| dP = 0$$

Proposición 2.2.6. *Si una sucesión de variables aleatorias* $\{X_n\}$ *es acotada en* $L^p(P)$*, entonces* $\{X_n\}$ *es uniformente integrable*[31].

2.2.2. Teoría asintótica

En esta sección se presenta algunos resultados de la teoría asintótica para sucesiones de variables aleatorias que serán de utilidad para derivar resultados análiticos posteriores.

[27] Véase Cinlar (2011) pag-189.
[28] Véase Bierens (2004) p-157
[29] Véase Bierens (1994) p-39
[30] Ver Billingsley (1995) p-338.
[31] Para una demostración de la proposición véase Gut (2013) pag-214.

2.2.2.1. Teoremas de convergencia para sucesiones de variables aleatorias

Definición 2.2.15. Convergencia casi segura. Sea $\{X_n\}$ una sucesión de variables aleatorias en el espacio de probabilidad $\big(\Omega, X, P\big)$. La sucesión de variables aleatorias $\{X_n\}$ converge de forma casi segura[32] o con probabilidad 1 a una variable aleatoria X, si y solo si:

$$\Pr\Big(\big\{\omega : X_n(\omega) \longrightarrow X(\omega)\big\}\, cuando\, n \to \infty\Big) \tag{2.2.2}$$

Y se denota como: $X_n \xrightarrow{a.s} X$.

Definición 2.2.16. Convergencia en probabilidad. Sea $\{X_n\}$ una sucesión de variables aleatorias en el espacio de probabilidad $\big(\Omega, X, P\big)$. La sucesión de variables aleatorias$\{X_n\}$ converge en probabilidad[33] a la variable aleatoria X , si y solo si para cada $\epsilon > 0$:

$$\lim_{n \longrightarrow \infty} \Pr\big(\big|X_n - X\big| > \epsilon\big) = 0 \tag{2.2.3}$$

Y se denota como: $X_n \xrightarrow{P} X$ y alternativamente como $plim\, X_n = X$.

Proposición 2.2.7. *Sea $\{a_n\}$ una sucesión de números reales[34] tal que $\lim_{n \longrightarrow \infty} a_n = c$, entonces se puede probar que[35] $plim\, a_n = c$.*

Definición 2.2.17. Convergencia en distribución. Sea $\{X_n\}$ una sucesión de variables aleatorias en el espacio de probabilidad $\big(\Omega, \mathcal{F}, P\big)$. Una sucesión de variables aleatorias $\{X_n\}$ con funciones de distribución denotadas por$\big\{F_n\big(X\big)\big\}$ se dice que convergen en distribución[36] a una variable aleatoria X con función de distribución $F\big(X\big)$ si y solo si:

$$\lim_{n \longrightarrow \infty} F_n\big(X\big) = F\big(X\big) \tag{2.2.4}$$

Y se denota como: $X_n \xrightarrow{d} X$.

[32] Véase Bhattacharya (2007), pag-7.
[33] Mayores detalles en Billinsley (1995), p-70.
[34] $\{a_n\}$ puede verse como una sucesión de variables aleatorias degeneradas.
[35] Véase Polansky (2011) p-106 para mayores detalles.
[36] Ver Shiryaev (1996), p-253.

La notación $X_n \xrightarrow{d} X$ significa que para valores largos de n se puede aproximar la función de distribución de X_n con la función de distribución de X. La distribución de de X es la distribución límite o la distribución asintótica.

Proposición 2.2.8. *Relación entre los modos de convergencia. Sea $\{X_n\}$una sucesión de variables aleatorias en el espacio de probabilidad $\big(\Omega, \mathcal{F}, P\big)$ y sea X una variable aleatoria definida en el mismo espacio de probabilidad, entonces se cumple lo siguiente*[37]*:*

$$(i)\, X_n \xrightarrow{a.s} X \implies X_n \xrightarrow{p} X \tag{2.2.5}$$

$$(ii)\, X_n \xrightarrow{p} X \implies X_n \xrightarrow{d} X \tag{2.2.6}$$

Proposición 2.2.9. *Teorema de Slutsky.*[38] *Sea $\{X_n\}$ y $\{Y_n\}$ dos sucesiones de variables aleatorias que se definen en un determinado espacio de probabilidad. Si $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{P} c$, para una constante $c \in \mathbb{R}$, entonces:*

$$(i)\, X_n + Y_n \xrightarrow{d} X + a \tag{2.2.7}$$

$$(ii)\, X_n Y_n \xrightarrow{d} aX \tag{2.2.8}$$

$$(iii)\, X_n / Y_n \xrightarrow{d} X/a, \;\; con\ a \neq 0 \tag{2.2.9}$$

Proposición 2.2.10. *Teorema de mapeo continuo. Sea $\{X_n\}$ una sucesión de variables aleatorias en el espacio de probabilidad $\big(\Omega, \mathcal{F}, P\big)$, tal que $X_n \xrightarrow{d} X$. Y sea $g\,(.)$ una función continua, entonces*[39]*:*

$$g\big(X_n\big) \xrightarrow{d} g\big(X\big) \quad cuando \quad n \to \infty \tag{2.2.10}$$

[37] La demostración de los resultados se encuentra en Gut (2013) p-209
[38] Una demostración de este teorema se puede encontrar en Shao (2003), p-60.
[39] Una demostración de este teorema se encuentra en Gut (2013), p-245.

2.2.3. Teorema Central del Límite y Ley Fuerte de los Grandes Números

2.2.3.1. Teorema del Límite Central

El **teorema central del límite** (CLT) es un resultado analítico de la estadística, que según determinadas condiciones la suma de una sucesión de variables aleatorias va a converger en distribución a otra variable aleatoria. Dentro de la teoría de probabilidad no existe un único CLT sino varios de ellos, y cada uno corresponde a las propiedades particuales presentes dentro de cada sucesión de variables aleatorias. A modo de ejemplo se pone el siguiente CLT:

Proposición 2.2.11. *CLT para sucesiones de variables aleatorias I.I.D*[40]. *Sea* $\{X_n\}$ *una sucesión de variables aleatorias independientes e identicamente distribuidas con expectativa* $E\big[X_1\big] = \mu$ *y varianza* $Var\big[X_1\big] = \sigma^2$. *Y sea* $S_n = X_1 + X_2 + \cdots + X_n$, $n \geq 1$. *Entonces:*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N\big(0,1\big) \tag{2.2.11}$$

2.2.3.2. Teorema del Límite Central Funcional

El **Teorema del Límite Central Funcional** (FCLT) es uno de varios principios de invarianza[41] desarrollados en la teoría de la probabilidad. A diferencia del CLT tradicional que se enfoca en la distribución límite de una unica suma S_n de variables aleatorias, el Teorema del Límite Central Funcional se enfoca en las propiedades límite de sucesiones de funciones aleatorias sobre espacios métricos, o sea $h(S_1, S_2, \ldots, S_n)$ donde S_n es una función aleatoria $\forall\, n$. En consecuencia, el FCLT es una generalización del CLT tradicional y de la misma manera que el CLT tradicional, no existe un único FCLT sino varios, cada uno va a depender del tipo de

[40] Una demostración del teorema se halla en Gut (2013), p-330.

[41] Un principio de invarianza, en un contexto de problema de límite probabilistico, se mantiene si la distribución límite existe y es independiente de las distribuciones particulares de las variables aleatorias envueltas, ver Fischer (2011), p -338.

proceso y sus propiedades probabilísticas. A modo de ejemplo se pone el siguiente resultado:

Proposición 2.2.12. *FCLT para variables I.I.D .*[42] *Sea*$\{\xi_n\}$ *una sucesión de variables aleatorias I.I.D, en el espacio de probabilidad* $(\Omega, \mathcal{F}, P)$*, con* $E[\xi_n] = 0$, $E[\xi_n^2] = \sigma^2$ *Y sea:* $S_n = \xi_1 + \xi_2 + \xi_3 + \cdots + \xi_n$, $n \in \mathbb{N}$. *Entonces para las funciones aleatorias* $X_n(t, w)$ *sobre el espacio* $\mathcal{C}[0,1]$ *definidas como:*

$$X_n(t,w) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}(w) + (nt - [nt]) \frac{1}{\sigma\sqrt{n}} \xi_{[nt]+1}(w) \tag{2.2.12}$$

Se cumple que $X_n \xrightarrow{d} W(t)$, donde $W(t)$ es un proceso de Wiener.

2.2.3.3. Ley fuerte de los grandes números

La ley fuerte de los grandes números (SLLN) es un resultado analítico de la estadística, que establece la convergencia casi segura de la suma de una sucesión de variables aleatorias hacia la expectativa de la misma. En teoría de la probabilidad no existe una única SLLN sino varias, y cada una corresponde a un tipo particular de variable aleatoria. A modo de ejemplo se pone la siguiente SLLN:

Proposición 2.2.13. *SLLN para sucesiones de variables aleatorias I.I.D*[43]. *Sea* $\{X_n\}$ *una sucesión de variables aleatorias independientes e identicamente distribuidas con expectativa* $E\big[X_1\big] = \mu$ *y* $E\big[X_1\big] < \infty$. *Y sea* $S_n = X_1 + X_2 + \cdots + X_n$, $n \geq 1$. *Entonces:*

$$\frac{S_n}{n} \xrightarrow{a.s} \mu$$

Corolario 2.2.1. *Sea* $\{X_n\}$ *una sucesión de variables aleatorias independientes e identicamente distribuidas con expectativa* $E\big[X_1^2\big] = \mu^2$ *y* $E\big[X_1^2\big] < \infty$. *Y sea*

[42] También llamado teorema de Donsker, una demostración del teorema esta en Cappaso y Bakstein (2012) p-307.

[43] Este resultado es conocido como la ley fuerte de Kolmogorov, una demostración del resultado se halla en Gut (2012), p-296.

$S_n^2 = X_1^2 + X_2^2 + \cdots + X_n^2\ ,\ n \geq 1$. *Entonces:*

$$\frac{S_n^2}{n} \xrightarrow{a.s} \mu^2$$

El corolario 2.2.1 implica que las leyes de los grandes números pueden extederse a momentos de orden superior si cumplen los supuestos respectivos.

Proposición 2.2.14. *Sea* $\{X_n\}$ *una sucesión de variables aleatorias tal que* $S_n = X_1 + X_2 + \cdots + X_n\ ,\ n \geq 1$ *, y sea* $\bar{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} E\big[X_i\big]$ *tal que* $\lim\limits_{n \longrightarrow \infty} \bar{\mu}_n = \mu$ *. Entoces se cumple que*[44]*:*

$$\frac{S_n}{n} - \bar{\mu}_n \xrightarrow{a.s} 0 \implies \frac{S_n}{n} \xrightarrow{a.s} \mu$$

2.2.4. Inferencia estadistica

En esta subsección se van a detallar conceptos básicos de la estadistica inferencial y definir algunos tipos de estimadores, dentro una gran variaedad de estos, que son necesarios para posteriores desarrollos en el presente trabajo.

Definición 2.2.18. Proceso Generador de datos (D.G.P)[45]. Es el verdado modelo estadistico subyacente que genera las observaciones en la muestra.[46]

Un ejemplo conocido es el lanzamiento de nueva moneda no trucada n veces, las n observaciones son generadas por el verdadero modelo estadistico que corresponde a una distribucion binomial.

Definición 2.2.19. Parámetro. Dado un modelo estadístico con función de densidad $f(x,\theta)$ un parámetro es un constante que determina una característica de la función de función de densidad tales como forma, ubicación, escala, proporción, entre otros[47] y se denota mediante θ.

[44]Véase Mittelhammer (2013), pag-260.
[45]Tambien llamado mecanismo generador de datos
[46]Para mayores detalles y su lugar en el proceso de inferencia,ver Lindsey (1996).
[47]Véase Spanos (1999), p-37.

Definición 2.2.20. Estimador y estadístico. Un estadistico $T = g(X_1, X_2, \ldots, X_n)$ se define como una función de las variables aleatorias contenidas en una muestra, y se denota como $T(X)$. Mientras que un *estimador*[48] es un estadistico $T(X)$ que se utiliza para estimar alguna cantidad poblacional θ, y se denota como $\hat{\theta}$.

De modo que un mismo estadístico puede ser usado por dos estimadores para hallar dos cantidades poblacionales distintas, por ejemplo: la media muestral $\bar{x}$ puede ser un estimador de la media poblacional y de la moda.

Definición 2.2.21. Estimador insesgado. Sea $\hat{\theta}$ un estimador de θ , se dice que $\hat{\theta}_n$ es un estimador insesgado[49] de θ si:

$$E\big[\hat{\theta}\big] = \theta$$

Definición 2.2.22. Estimador eficiente. Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados de θ. Se dice que $\hat{\theta}_1$ es es eficiente[50], también llamado relativamente eficiente, con respecto a $\hat{\theta}_2$ si se cumple:

$$Var\big(\hat{\theta}_1\big) \leq Var\big(\hat{\theta}_2\big)$$

Definición 2.2.23. Estimador consistente. Un Estimador $\hat{\theta}_n$ de θ es consistente[51] si para cualquier número positivo ε, si se cumple:

$$\lim_{n \longrightarrow \infty} \Pr\big(\big|\hat{\theta}_n - \theta\big| > \epsilon\big) = 0$$

O de manera equivalente:

$$\lim_{n \longrightarrow \infty} \Pr\big(\big|\hat{\theta}_n - \theta\big| \leq \epsilon\big) = 1$$

[48] Mayores detalles en Panik (2005), pag-373.
[49] Véase Westfall (2013), pag-284.
[50] Mayores detalles en Liero (2012), pag-97.
[51] Mayores detalles en Panik (2005), pag-411.

Definición 2.2.24. Estimador asintóticamente normal. Un Estimador $\hat{\theta}_n$ de θ , es asintóticamente normal[52] si se puede encontrar una sucesión normalizadora $\{X_n\}$ tal que:

$$C_n\big(\hat{\theta}_n - \theta\big) \ \widetilde{a} \ N\big(0, Var(\hat{\theta}_n)\big)$$

2.3. Procesos estocásticos y tipos de procesos estocásticos

Definición 2.3.1. Proceso estocástico. Sea $\big(\Omega, \mathcal{F}, P\big)$ un espacio de probabilidad, $\big(X, \mathfrak{B}\big)$ un espacio medible y T un conjunto de índices. Un **proceso estocástico**[53] $(X_t)_{t\in T}$ es una familia o conjunto de variables aleatorias en un espacio de probabilidad común, tal que: $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ para $t \in T$, y se denota como $\{X_t\}$.

Por definición un proceso estocástico, se compone de un espacio muestral, funciones indexadas temporalmente y una medida de probabilidad,[54] de modo que un proceso estocástico es una función que asigna a cada elemento del espacio muestral una trayectoria temporal asociada a una distribución de probabilidad; y en general un proceso estocástico $\{X_t\}$ tiene momentos probabilísticos (por ejemplo: media y varianza) que son funciones del tiempo.

2.3.1. Tipos de procesos estocásticos

Existen diversas formas de clasificar los procesos estocásticos según el tipo de propiedades estadísticas y probabilísticas que posean. En el presente trabajo los diversos desarrollos analíticos y teoremas requieren un uso intensivo de los siguientes tipos de procesos estocásticos:

2.3.1.1. Procesos dependientes e independientes

[52] Véase Jiang (2010), pag-89.
[53] Ver Capasso (2012) para mayores detalles sobre los procesos estocásticos.
[54] Ver Venkatarama (2006) p-406, para mayores detalles y ejemplos de procesos estocásticos.

Un proceso estocástico $\{X_t\}$ es independiente[55] si se cumple lo siguiente:

$$F_{X_1X_2\ldots X_n}(X_1X_2\ldots X_n)=F_{x_1}(X_1,t_1)\times F_{x_2}(X_2,t_2)\ldots\times F_{x_n}(X_n,t_n)=\prod_{i=1}^{n}F_{x_i}(X_i,t_i)$$

Lo cual quiere decir que la distribución conjunta del proceso estocástico $\{X_t\}$ para t instantes temporales $(t_1,t_2,\ldots t_n)$ se puede expresar como el producto de la función de distribución mismo proceso para los n instantes del tiempo. En caso contrario el proceso $\{X_t\}$se dice que es dependiente.

2.3.1.2. Procesos estacionarios y no estacionarios

Un proceso estocástico $\{X_t\}$ es estacionario en sentido estricto[56], también llamado estrictamente estacionario, si su distribución conjunta en el intervalo temporal $\{t_i : i=1,2,...n\}$ es la misma que en el intervalo $\{t_i+h \quad : i=1,2,...n\}$ para todo h. En otras palabras:

$$F_{X_1X_2\cdots X_n}(X_1X_2\cdots X_n)=F_{X_{1+h}X_{2+h}\cdots X_{n+h}}(X_1X_2\cdots X_n)$$

De lo cual se desprende que la función de distribución un proceso estocástico estricamente estacionario $X(t)$ se mantiene invariante ante cualquier desplazamiento temporal. Mientras que un proceso estocástico $X(t)$ es estacionario de segundo orden[57], también llamado estacionario en covarianza o débilmente estacionario, si se cumple que:

1. $E\left[X(t)\right]=\mu_x$para todo $-\infty<t<+\infty$

2. $E\left[X(t)^2\right]<\infty$ para todo $-\infty<t<+\infty$

3. $E\left[\left(X(s)-\mu\right)\left(X\left(t\right)-\mu\right)\right]$ depende de la diferencia temporal $|s-t|$

[55] Véase Kobayashi (2012), p-318.
[56] Mayores detalles en Doob (1953), pag-94.
[57] Véase Karlin (1975), pag-445.

De modo que un proceso estacionario $X(t)$ es un proceso estocástico cuyas leyes probabilísticas no cambian por desplazamientos temporales. En otras palabras un proceso débilmente estacionario $X(t)$ tiene momentos de primer y segundo orden que no son funciones del tiempo.

Un proceso estocástico $X(t)$ es no estacionario cuando sus leyes probabilisticas no satisfacen los supuestos necesarios para ser estacionario, en otras palabras, las leyes probabilisticas de un proceso no estacionario son funciones del tiempo. Un ejemplo conocido es la caminata aleataria.

2.3.1.3. Procesos de Wiener

Un proceso de Wiener[58], también llamado proceso Browniano o movimiento Browniano, es un proceso estocástico de tiempo continuo que satisface las siguientes propiedades:

- Homogeneidad espacial:

$$\Pr\left[W\left(t\right) \leq w \mid W\left(0\right) = a\right] = \Pr\left[W\left(t\right) \leq w + b \mid W\left(0\right) = a + b\right]$$

- Homogeneidad temporal:

$$\Pr\left[W\left(t\right) \leq w \mid W\left(0\right) = a\right] = \Pr\left[W\left(t+s\right) \leq w \mid W\left(s\right) = a\right]$$

- Incrementos independientes:

  Para un conjunto de intervalos disjuntos $\left(s_i, t_i\right], i = 1, 2, \ldots$ se tiene que:

$$W\left(t_i\right) - W\left(t_s\right), i = 1, 2, \ldots$$

  Son incrementos independientes.

- Propiedad Gausiana:

[58] Véase Kobayashi (2012), p-491. Para extenso análisis de los procesos de Wiener y sus propiedades véase Borodin y Salmien (2002) .

Cualquier incremento $W(t_i) - W(s_i), t_i > s_i$ es normalmente distribuido[59]:

$$\Pr\left[W(t) \leq x \mid W(t_0) = x_0\right] = \frac{1}{\sqrt{2\pi\alpha(t-t_0)}} \int_{-\infty}^{x-x_0} \exp\left\{-\frac{y^2}{2\alpha(t-t_0)}\right\} dy$$

- Propiedad de Markov:

$$\Pr\left[W(t+s) \leq w \mid W(u), u \leq t\right] = \Pr\left[W(t+s) \leq w \mid W(t)\right], \forall s \geq 0$$

2.4. Procesos Strong-Mixing

En probabilidad dos eventos A y B son independientes[60] en un mismo espacio de probabilidad cuando:

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B)$$

Y son dependientes cuando $\Pr(A \cap B) - \Pr(A) \times \Pr(B) \neq 0$. De la misma manera dos eventos generados por los sub-espeacios de probabilidad $\mathcal{A}$ y $\mathcal{B}$, son independientes si se cumple:

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B) \quad \textit{con } A \in \mathcal{A} \textit{ y } B \in \mathcal{B}$$

De modo que una medida útil para cuantificar la dependencia entre dos eventos es la siguiente:

$$\left|\Pr(A \cap B) - \Pr(A) \times \Pr(B)\right|$$

Definición 2.4.1. Dependencia α-mixing[61]. Sea $\mathcal{A}$ y $\mathcal{B}$, dos sub-σ álgebras de $\mathcal{F}$, entonces la medida de dependenciaα-mixing se define :

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} \left|\Pr(A \cap B) - \Pr(A) \times \Pr(B)\right|$$

[59] Para mayores detalles sobre la distribución normal véase el apéndice (A.4)

[60] En probabilidad la independencia se da a nivel eventos, variables aleatorias, espacios y otras estructuras probabilísticas.

[61] Véase Bradley (2005).

Definición 2.4.2. Coeficiente α-mixing. Sea $\mathcal{F}_{-\infty}^{t} = \sigma\big(X_t, X_{t-1}, X_{t-2}, \ldots\big)$ y $\mathcal{F}_{t}^{\infty} = \sigma\big(X_t, X_{t+1}, X_{t+2}, \ldots\big)$ un par de σ- álgebras generados por$\{X_t\}$. Un coeficiente α-mixing[62] se define como:

$$\alpha\left(t\right) = \sup_{t} \alpha\left(\mathcal{F}_{1}^{t}, \mathcal{F}_{t+n}^{\infty}\right)$$

Definición 2.4.3. Procesos α-mixing[63]. Sea $\mathcal{F}_{-\infty}^{t} = \sigma\big(X_t, X_{t-1}, X_{t-2}, \ldots\big)$ y $\mathcal{F}_{t}^{\infty} = \sigma\big(X_t, X_{t+1}, X_{t+2}, \ldots\big)$, un proceso $X(t)$ es α-mixing o strong-mixing si:

$$\alpha\left(t\right) = \sup_{t} \alpha\left(\mathcal{F}_{1}^{t}, \mathcal{F}_{t+n}^{\infty}\right) \longrightarrow 0 \quad cuando \quad t \to \infty \tag{2.4.1}$$

O equivalentemente:

$$\alpha\left(t\right) = \sup_{t} \sup_{A \in \mathcal{F}_{-\infty}^{t}, B \in \mathcal{F}_{t}^{\infty}} \Big|\Pr(A \cap B) - \Pr(A) \times \Pr(B)\Big| \longrightarrow 0 \quad cuando \quad t \to \infty$$

De la definición (2.4.3) se tiene que un proceso estocástico $\{X_t\}$ es strong mixing si la sucesión de coeficientes α-mixing tiende a 0, y se desprende que a medida que el coeficiente α-mixing tiende a 0, la dependencia dentro del proceso va disminuyendo y en el límite el proceso $\{X_t\}$ es asintóticamente independiente.

Una cualidad resaltante de los procesos strong-mixing es que pueden ser tanto procesos estacionarios como no estacionarios, mientras que otros procesos asintóticamente independientes como los procesos ergódicos[64] solo pueden ser estacionarios, de modo que los procesos strong mixing son más generales que los procesos ergódicos.

Definición 2.4.4. Tamaño de una sucesión[65]. Sea $\alpha\left(t\right)$ una sucesión tal que $\alpha\left(t\right) = O\left(m^{-a-\varepsilon}\right)$ para algún $\varepsilon > 0$, entonces $\alpha\left(t\right)$ es de tamaño $-a$.

[62] Ver Gut (2013) p-450.

[63] Mayores detalles en Zhengyan (1996) p-4.

[64] Un proceso $\{Z_t\}$ es ergódico si para cualquier par de funciones acotadas $f : \mathbb{R}^k \longrightarrow \mathbb{R}$ y $g : \mathbb{R}^l \longrightarrow \mathbb{R}$, se tiene: $\lim_{n \longrightarrow \infty} \Big|E\Big[f\big(z_i, \ldots z_{i+k}\big) g\big(z_{i+n}, \ldots z_{i+n+l}\big)\Big]\Big| = \Big|E\Big[f\big(z_i, \ldots z_{i+k}\big)\Big]\Big| \Big|E\Big[g\big(z_{i+n}, \ldots z_{i+n+l}\big)\Big]\Big|$.

[65] Véase White (2000) p-49.

Dado que el coeficiente α-mixing tiende a 0 se puede interpretar como una sucesión de coeficientes α-mixing que tiene una tasa de convergencia. Si la sucesión de coefientes α-mixing es a lo mucho de orden $t^{-a-\epsilon}$ $\epsilon > 0$, denotado como $\alpha(t) = O\big(t^{-a-\epsilon}\big)$ $\epsilon > 0$, entonces se dice que α es de tamaño $-a$.

Proposición 2.4.1. *Sea $X_t = g\big(Z_t, Z_{t-1}, \ldots, Z_{t-\tau}\big)$ donde $g\,(.)$ una función medible y τ es un entero positivo. Si Z_t es $\alpha-$mixing con coeficiente mixing $-a$, entonces $\big\{X_t\big\}$ sigue un proceso $\alpha-$mixing con coeficiente mixing $-a$.*[66]

Y generalización para vectores de variables aleatorias es la siguiente:

Proposición 2.4.2. *Sea el vector de procesos estocásticos $\{\mathbf{X_t}, \mathbf{y_t}\}$, donde $\mathbf{X_t}$ es un vector de procesos estocásticos y $\mathbf{y_t}$ es un proceso estocástico, α-mixing con tamaño $-a$, entonces $\big\{\mathbf{X_t}\mathbf{X_t'}\big\}$, $\{\mathbf{X_t}\mathbf{y_t}\}$ y $\big\{\mathbf{y_t}'\mathbf{y_t}\big\}$ es un vector de procesos estocásticos α-mixing con tamaño $-a$*[67].

Un resultado importante para desarrollos posteriores es una SLLN que corresponde a procesos strong mixing y que se presenta a continuación:

Teorema 2.4.1. *SLLN para sucesiones $\alpha-$mixing.*[68] *Sea $\big\{X_t\big\}$ una sucesión de variables aleatorias $\alpha-$mixing , con coeficiente α de tamaño $r/(r-1)$, $r > 1$ y $\mu_t \equiv E\big[X_t\big]$, tal que $E\big|X_t\big|^{r+\delta} \leq k < \infty$,$\delta > 0\,\forall t$. Y sea $S_n = X_1 + X_2 + \cdots + X_n$, $n \geq 1$. Entonces:*

$$\frac{S_n}{n} - \bar{\mu}_n \xrightarrow{a.s} 0 \tag{2.4.2}$$

Donde $\bar{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} E\big[X_i\big]$.

Teorema 2.4.2. *FCLT para sucesiones de variables aleatorias α-mixing. Sea $\{X_n\}$ una sucesión de variables aleatorias en el espacio de probabilidad $\big(\Omega, X, P\big)$, y sea*

[66] Ver lema 2.1 de White and Domowitz (1984).
[67] Véase White (1999) p-50.
[68] El teorema se presenta como corolario 3.48 en White (2000) y esta basado en el teorema 2.10 de McLeish (1975).

$S_n = \sum_{i=1}^{n} X_i$ *con* $n \in \mathbb{N}$*, que satisface:*

$$
\begin{array}{ll}
(i)\, E\big[X_n\big] = 0 & ,\, n \in \mathbb{N} \\
(ii)\, E\big[X_n^2\big] < \infty & ,\, n \in \mathbb{N} \\
(iii)\, \lim\limits_{n \longrightarrow \infty} \dfrac{E\big[S_n^2\big]}{n} = \sigma^2 & ,\, \sigma > 0 \\
(iv)\, \limsup E^{1/\beta}\big|X_n\big|^{\beta} < \infty & ,\, \beta \in \big(2, \infty\big] \\
(v)\, \sum\limits_{i \in \mathbb{N}} \alpha(i)^{1-\gamma} < \infty & ,\, \gamma = 2/\beta
\end{array}
$$

Y se define la función aleatoria $W_n : \Omega \longrightarrow \mathcal{D}$ como: $W_n(t) = {}^{S_{[nt]}}/_{\sigma\sqrt{n}}\, , t \in [0, 1]$. Entonces:

$$W_n(t) = \frac{S_{[nt]}}{\sigma\sqrt{n}} \xrightarrow{d} W\left(t\right) \tag{2.4.3}$$

Siendo $W\left(t\right)$ un proceso de Wiener standar[69].

Corolario 2.4.1. *Si* $t = 1$ *entonces* $W_n(1) = S_n/\sigma\sqrt{n} \xrightarrow{d} W\left(1\right) = N\left(0, 1\right)$[70].

Proposición 2.4.3. *Se cumple que* $\int_0^1 sdW(s) \sim N\left(0, 1/3\right)$[71].

[69] La demostración del teorema se encuentra en Herrdorf (1984), colorario 1, p -142)
[70] Véase Hassler (2106), p-307.
[71] Véase Ibid p-203, para unda demostración de este resultado

Capítulo 3

Regresion Lineal Múltiple

3.1. Modelo de regresión lineal

Una regresión[1] es un método estadístico cuyo propósito es determinar si existen relaciones entre la variable dependiente y variables independientes, y al mismo tiempo cuantificar el impacto de las variables explicativas sobre la variable explicada. Tal relación toma la forma de una función lineal con un término de error aditivo. Una regresión es lineal en los parámetros (y de ahora en adelante: regressión lineal) cuando los parametros tienen formas funcionales lineales y es múltiple cuando existe más de una variable explicativa y una sola variable explicada (que contrasta con las regresiones multivariadas donde existen varias variables dependientes y varias variables independientes). En consecuencia una regresión lineal multiple[2] se puede expresar como:

$$
\begin{aligned}
Y_1 &= \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_k X_{1k} + \varepsilon_1 \\
Y_2 &= \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_k X_{2k} + \varepsilon_2 \\
\vdots &= \quad \vdots \quad + \quad \vdots \quad + \cdots + \quad \ddots \quad + \vdots \\
Y_T &= \beta_1 X_{T1} + \beta_2 X_{T2} + \cdots + \beta_k X_{Tk} + \varepsilon_T
\end{aligned}
$$

[1] también llamado modelo de regresión o analisis de regresion

[2] Una excelente introducción a los modelos de regresión lineal se puede encontrar en Seber y Lee (2003).

O en forma vectorial como:

$$y_t = \mathbf{x}_t'\beta + \varepsilon_t \qquad (t = 1, 2, \ldots, T)$$

Y en forma matricial:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} \\ X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{T1} & X_{T2} & \cdots & X_{Tk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix} \tag{3.1.1}$$

Y reescribirse como:

$$\mathbf{y} = \mathbf{X}'\beta + \varepsilon \tag{3.1.2}$$

Donde $\mathbf{y}$ es un vector de observaciones de orden $T \times 1$, $\mathbf{X}$ es una matrix orden $T \times k$ conocida como matriz de observaciones,β es un vector fila de orden $T \times 1$ que contiene k parametros poblacionales desconocidos y ε es un vector fila de orden $T \times 1$ que contiene T perturbaciones.

3.2. Método de los MCO en regresiones lineales.

El método de los Mínimos Cuadrados Ordinarios (MCO) es un método de optimización matemático, basado en el principio de minimos cuadrados, que se usa para hallar la mejor aproximación a la solución de un sistema de ecuaciones. En estadistica el método MCO es uno de varios métodos utilizados para hallar los estimadores en un modelo de regresión.[3] En el caso de una regresion lineal el método MCO minimiza la suma de los cuadrados de los residuos, que se definen como:

$$e_t = y_t - \hat{y}_t = y_t - \mathbf{x_t'}\tilde{\beta}$$

[3]Otros métodos son: Método de Máxima Verosimilitud, Método General de los Momentos, Método Bayesiano, etc. Véase Golberg y Cho (2003) para la aplicación de éstos métodos en el modelo de regresión lineal.

Con $\widetilde{\beta}$ un valor hipotético de β

El término e_t es el residuo de la observación i, que se expresa como la diferencia entre el valor actual de y_t y el valor de Y_t predicho que se obtiene al reemplazar $\tilde{\beta}$ por β , en $\mathbf{x_t'}\beta$. De modo que la expresión:

$$\sum_{i=1}^{T} e_i^2 = \sum_{i=1}^{T} \left(y_{t-}\mathbf{x_t'}\widetilde{\beta}\right)^2 = \mathbf{e'e} = \left(\mathbf{y} - \mathbf{X'}\widetilde{\beta}\right)' \left(\mathbf{y} - \mathbf{X'}\widetilde{\beta}\right)$$

Es la suma de los cuadrados de los residuos (SCR), de la expresión anterior se desprende que la SCR es una función de $\tilde{\beta}$ porque los residuos de cada observación dependen de $\tilde{\beta}$. Entonces el estimador MCO $\mathbf{b}$ de β, es el que minimiza la siguiente función:

$$\mathbf{b} \equiv \arg\min_{\widetilde{\beta}} SSR(\tilde{\beta}) = \left(\mathbf{y} - \mathbf{X'}\widetilde{\beta}\right)' \left(\mathbf{y} - \mathbf{X'}\widetilde{\beta}\right)$$

Y mediante manipulaciones algebraicas[4]se obtiene que[5] :

$$\mathbf{b} = \left[\mathbf{X'X}\right]^{-1}\mathbf{X}\text{`}\mathbf{y} \tag{3.2.1}$$

A modo de ejemplo sea $y_t = \alpha + \beta_1 x_t + \varepsilon_t$ el modelo de regresión con intercepto y una variable independiente x_t, entonces el vector de estimadores $\mathbf{b}$ de la ecuación (3.2.1) toma la siguiente forma:

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \end{bmatrix} = \left[\begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_T \end{pmatrix} \times \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_T \end{pmatrix} \right]^{-1} \left[\begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_T \end{pmatrix} \times \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} \right]$$

y queda como:

[4]El único supuesto necesario para hallar el vector de estimadores $\mathbf{b}$ es que la matriz $\mathbf{X}$ sea de rango completo, lo cual permite que $\left(\mathbf{X'X}\right)^{-1}$ sea invertible.

[5]Una completa explicación de la derivación de los resultados presentados en esta sección se puede obtener en Greene (2012) p-26.

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} x_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} y_t \\ \sum_{t=1}^{T} x_t y_t \end{bmatrix} \tag{3.2.2}$$

De manera alternativa se puede hallar el vector de estimadores $(\mathbf{b} - \beta)$ para el modelo de regresión. El punto de partida es la de la ecuación (3.2.1), tal que:

$$\mathbf{b} = \left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}`\mathbf{y}$$

$$\mathbf{b} = \left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}`\left[\mathbf{X}'\beta + \varepsilon\right], \qquad \textit{dado que } \mathbf{y} = \mathbf{X}'\beta + \varepsilon$$

$$\mathbf{b} = \beta + \left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}`\varepsilon$$

$$\mathbf{b} - \beta = \left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}`\varepsilon$$

De modo que:

$$(\mathbf{b} - \beta) = \left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}`\varepsilon \tag{3.2.3}$$

Y alternativamente:

$$(\mathbf{b} - \beta) = \left[\sum_{i=1}^{T} \mathbf{x}_i\mathbf{x}_i'\right]^{-1} \left[\sum_{i=1}^{T} \mathbf{x}_i\varepsilon_i\right] \tag{3.2.4}$$

Cuando $y_t = \alpha + \beta_1 x_t + \varepsilon_t$, la ecuación (3.2.3)se expresa como:

$$\begin{bmatrix} \hat{\alpha} - \alpha \\ \hat{\beta}_1 - \beta_1 \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} x_t^2 \end{bmatrix}^{-1} \times \begin{bmatrix} \sum_{t=1}^{T} \varepsilon_t \\ \sum_{t=1}^{T} x_t\varepsilon_t \end{bmatrix} \tag{3.2.5}$$

En este punto tiene que quedar claro que el método MCO halla el estimador $\mathbf{b}$ del vector de parametros β en la regresión (3.1.2), sin embargo no determina ninguna de las propiedades estadísticas que pueda tener $\mathbf{b}$. Los estimadores MCO, al igual que cualquier estimador de un parámetro poblacional, deben cumplir ciertas propiedades estadisticas para ser preferidos sobre otros estimadores. Estas propiedades, conocidas como propiedades deseables de un estimador, son: insesgadez, eficiencia, consistencia y normalidad asintótica.

3.3. Propiedades estadísticas de los estimadores MCO en regresiones lineales múltiples.

Las propiedades estadísticas de los estimadores MCO en regresiones lineales van a depender principalmente de 3 aspectos: Primero, el tipo de modelo de regresión lineal bajo consideración, lo que significa si una regresión contiene variables dinámicas, estáticas o una combinación de ambas. Segundo, la naturaleza de la matriz de datos $\mathbf{X}$, en el sentido de que si las observaciones provienen de variables estocásticas o determinísticas, y al mismo tiempo si son observaciones de corte transversal, series temporales o datos de panel. Tercero, la relación conjunta entre las variables en $\mathbf{X}$ con las pertubaciones ε, de modo que se pueda asumir que son independientes, no correlacionadas o la existencia de momentos conjuntos igual a 0, entre otros.

Las propiedades de los estimadores MCO se pueden dividir en dos grupos: propiedades en muestra pequeña, que son: insesgadez y eficiencia; y las propiedades en muestra grande: consistencia y normalidad asintótica. Resulta claro del párrafo anterior que dependiendo de cómo se especifiquen los 3 aspectos mencionados, los estimadores MCO, $\mathbf{b}$ de β , en una regresión, pueden tener todas las propiedades deseables como no pueden tener ninguna, o algunas de ellas. A continuación se enumeran un par de resultados a modo de ejemplo.

Proposición 3.3.1. *Propiedades de los estimadores MCO con regresores fijos. Sea* $\mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}`\mathbf{y}$ *el estimador MCO de* β *, entonces bajo los siguientes supuestos:* (i) $E\left[\mathbf{Y}\right] = \mathbf{X}\beta$, (ii) $\mathbf{X}_{T\times k}$ *una matriz de regresores fijos con rango* k, $(iii)\,\varepsilon \sim i.i.d$

con $E\big[\varepsilon\big] = \mathbf{0}$ *y* $Cov\big[\varepsilon\big] = \sigma^2\mathbf{I}$, *(iv)* $\lim_{n\to\infty}\big(T^{-1}\mathbf{X}'\mathbf{X}\big) = \mathbf{Q}$*, se comprueba*[6] *que:*

$$(i)\, E\big[\mathbf{b}\big] = \beta \qquad insesgadez$$

$$(ii)\, Var\big[\mathbf{b}\big] \leq Var\big[\mathbf{b}'\big] \qquad eficiencia$$

$$(iii)\, plim\, \mathbf{b} = \beta \qquad consistencia$$

$$(iii)\, \sqrt{T}(\mathbf{b}-\beta) \xrightarrow{\ d\ } N(\mathbf{0}, \sigma^2\mathbf{Q}^{-1}) normalidad\ asintótica$$

Proposición 3.3.2. *Propiedades de los estimadores MCO con regresores predeterminados. Sea* $\mathbf{b} = \big(\mathbf{X}'\mathbf{X}\big)^{-1}\mathbf{X}`\mathbf{y}$ *el estimador MCO de* β *, entonces bajo los siguientes supuestos: (i)* $\big\{y_t, \mathbf{x}_t\big\}$ *es un proceso estocástico conjuntamente estacionario y ergódico, (ii)* $\mathbf{x}_t$ *es un vector de regresores predeterminados, esto es:* $E\big[\mathbf{x}_t \cdot \big(y_t - \mathbf{x}_t'\beta\big)\big] = 0$ *. (ii)* $E\big[\mathbf{x}_t\mathbf{x}_t'\big]$ *es una matriz finita y no singular de orden* $k \times k$*. Entonces:* $plim\, \mathbf{b} = \beta$[7].

Se evidencia que bajo los supuestos de la proposición 3.3.1, el estimador MCO **b** tiene todas las propiedades deseables de un estimador, mientras que el mismo estimador **b**, bajo los supuestos de la proposición 3.3.2, solamente es consistente. Esta diferencia es importante dado que en un contexto de series temporales, no se puede asumir que las observaciones son determinísticas sino que son realizaciones de procesos estocásticos, como el especificado en la proposición 3.3.2. En general, cuando los regresores son variables que siguen procesos dependientes el estimador **b** de β, va a ser sesgado e ineficiente debido a la relación de dependencia entre las observaciones en distintos períodos de modo que las únicas propiedades deseables de **b** que se pueden obtener son las de muestra larga.

[6] para la demostración de este teorema bajo estos supuestos y otros similares ver Mittelhammer (2013), cap. 8

[7] La demostración de este resultado se encuentra en Hayashi (2000), p-114.

3.4. Inferencia en el modelo de regresión lineal

Bajo los supuestos de la proposición (3.3.1), en especial la distribución de los estimadores $\hat{\beta}_j$, se pueden establecer los siguientes resultados inferenciales en el modelo de regresión lineal:

Proposición 3.4.1. *Prueba t en el modelo de regresión. La prueba de significancia individual para los estimadores $\hat{\beta}_j$ se realiza al contrastar la hipótesis nula $H_0 : \beta_j = \beta_j^0$ versus la hipótesis alternativa $H_A : \beta_j \neq \beta_j^0$se realiza mediante el estadístico:*

$$t_j = \frac{\hat{\beta}_j - \beta_j^0}{se\left(\hat{\beta}_j\right)}$$

y sigue una distribución $t \sim_{(n-k-1,\alpha/2)}$con $n-k-1$ grados de libertad y α es el nivel de confianza[8] .

Proposición 3.4.2. *Prueba F en el modelo de regresión. La prueba de significancia conjunta de los estimadores $\hat{\beta}_j$ se realiza al contrastar la hipotesis nula $H_0 : \beta_1 = \beta_2 = \beta_3 = \ldots \beta_k = 0$ versus la hipótesis alternativa usando el estadístico:*

$$F = \frac{R_k^2/k}{\left(1 - R_k^2\right)/n - k - 1}$$

donde R_k^2 es el coeficiente de correlación muestral múltiple del modelo de regresión con k regresores. Y estadístico F sigue una distribución $F \sim F_{k,k-n-1,\alpha}$ con k y $n-k-1$ grados de libertad y α es el nivel de confianza[9].

Proposición 3.4.3. *El intervalo de confianza del estimador $\hat{\beta}_j$ en el modelo de regresión lineal es $\hat{\beta}_j \pm t \sim_{(n-k-1,\alpha/2)} \times se\left(\hat{\beta}_j\right)$*[10].

Queda en relevancia que la distribución de los estimadores $\hat{\beta}_j$ influyen de manera determinante en los tests estadísticos y son crucial a la hora de realizar inferencias en el modelo de regresión lineal y establecer conclusiones estadísticas.

[8] Una demostración de este resultado se encuentra en Sen y Srivastava (1990) ,capítulo 3.
[9] Ibid
[10] Ibid

Capítulo 4

Propiedades asintóticas de los estimadores MCO en una regresión con tendencia y variables que siguen procesos dependientes strong mixing

En esta parte se van a determinar las propiedades asintóticas de los estimadores MCO en una regresión que tiene como regresores a variables que siguen procesos dependientes α-mixing, un componente tendencial y errores que también siguen procesos dependientes α-mixing. El propósito es determinar de manera teórica si la distribución del estimador $\mathbf{b}$ de β existe o no, y en caso de existir determinar si es igual a la distribución de los estimadores $\mathbf{b}$ de β cuando los regresores tienen otras propiedades estadístivas y probabilísticas.

4.1. Definición del modelo de regresión

La variable dependiente y_t sigue el siguiente proceso generador de datos (D.G.P):

$$y_t = \alpha + \delta t + \beta x_t + \varepsilon_t \qquad para\, t = 1, 2 \ldots, T \tag{4.1.1}$$

Donde: α es un término constante, t es un componente tendencial determinístico, x_t es una variable que sigue un proceso α-mixing y ε_t son las perturbaciones que también siguen un proceso α-mixing. Esta especificacion indica que la variable y_t tiene un componente tendencial autónomo y a la vez depende otra variable x_t. El modelo que se presenta en la ecuación (13.1) es bastante general, ya que al seguir x_t un proceso α-mixing permite incorporar variables estacionarias como no estacionarias y así como permite heterogeneidad en las distribuciones de las observaciones. Otras especificaciones de x_t como la de un un proceso estacionario o deterministico que no permite incorporar variables no estacionarias y heterogeneas. Al mismo tiempo, al especificar que las perturbaciones siguen procesos α-mixing se permite también que las perturbaciones esten correlacionadas y sean heterogeneas.

4.2. Derivación de la distribución asintótica

Para hallar la distribución asintótica del modelo (4.1.1) se aplica el método M.C.O para hallar el estimador $\mathbf{b}$. Dado que el modelo (4.1.1) se puede expresar como $\mathbf{y} = \mathbf{X}'\beta + \varepsilon$ con $\beta' = \begin{bmatrix}\alpha & \delta & \beta\end{bmatrix}$, el estimador $\mathbf{b}$ de β toma la forma de la ecuación (3.2.1):

$$\mathbf{b} = \left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}`\mathbf{y}$$

Y de manera análoga cuando $y_t = \alpha + \delta t + \beta x_t + \varepsilon_t$ la ecuación (3.2.3) para el vector $(\mathbf{b} - \beta)$ es:

$$\begin{bmatrix}\hat{\alpha} - \alpha \\ \hat{\delta} - \delta \\ \hat{\beta} - \beta\end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^{T} t & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 & \sum_{t=1}^{T} t x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} t x_t & \sum_{t=1}^{T} x_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} \varepsilon_t \\ \sum_{t=1}^{T} t\varepsilon_t \\ \sum_{t=1}^{T} x_t\varepsilon_t \end{bmatrix} \tag{4.2.1}$$

Ahora se procede a pre-multiplicar ambos lados de la ecuación matricial (4.2.1) por la siguiente matriz de escala[1]:

$$\mathit{\Gamma}_T = \begin{bmatrix} \sqrt{T} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & \sqrt{T} \end{bmatrix}$$

De modo que la ecuación (4.2.1)queda como:

$$\begin{bmatrix} \sqrt{T} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & \sqrt{T} \end{bmatrix} \times \begin{bmatrix} \hat{\alpha}-\alpha \\ \hat{\delta}-\delta \\ \hat{\beta}-\beta \end{bmatrix} = \begin{bmatrix} \sqrt{T} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & \sqrt{T} \end{bmatrix} \times \begin{bmatrix} T & \sum_{t=1}^{T} t & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 & \sum_{t=1}^{T} t x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} t x_t & \sum_{t=1}^{T} x_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} \varepsilon_t \\ \sum_{t=1}^{T} t\varepsilon_t \\ \sum_{t=1}^{T} x_t\varepsilon_t \end{bmatrix}$$

Y dado que $\mathit{\Gamma}_T \mathit{\Gamma}_T^{-1} = \mathbf{I}$

$$\begin{bmatrix} \sqrt{T}(\hat{\alpha}-\alpha) \\ T^{3/2}(\hat{\delta}-\delta) \\ \sqrt{T}(\hat{\beta}-\beta) \end{bmatrix} = \begin{bmatrix} \sqrt{T} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & \sqrt{T} \end{bmatrix} \begin{bmatrix} T & \sum_{t=1}^{T} t & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 & \sum_{t=1}^{T} t x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} t x_t & \sum_{t=1}^{T} x_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sqrt{T} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & \sqrt{T} \end{bmatrix}$$

$$\times \begin{bmatrix} \sqrt{T} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & \sqrt{T} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} \varepsilon_t \\ \sum_{t=1}^{T} t\varepsilon_t \\ \sum_{t=1}^{T} x_t\varepsilon_t \end{bmatrix}$$

[1]Se hace uso de matrices de escala porque las variables independientes tienen diferentes tasas de convergencia que implica que si se usa una escala única algunas variables divergen o colapsan.

$$
\begin{bmatrix} \sqrt{T}(\hat{\alpha}-\alpha) \\ T^{3/2}(\hat{\delta}-\delta) \\ \sqrt{T}(\hat{\beta}-\beta) \end{bmatrix} = \left\{ \begin{bmatrix} \sqrt{T} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & \sqrt{T} \end{bmatrix}^{-1} \begin{bmatrix} T & \sum_{t=1}^{T} t & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 & \sum_{t=1}^{T} t x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} t x_t & \sum_{t=1}^{T} x_t^2 \end{bmatrix} \begin{bmatrix} \sqrt{T} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & \sqrt{T} \end{bmatrix}^{-1} \right\}^{-1}
$$

$$
\times \left\{ \begin{bmatrix} \sqrt{T} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & \sqrt{T} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} \varepsilon_t \\ \sum_{t=1}^{T} t\varepsilon_t \\ \sum_{t=1}^{T} x_t\varepsilon_t \end{bmatrix} \right\}
$$

De modo que:

$$
\begin{bmatrix} \sqrt{T}(\hat{\alpha}-\alpha) \\ T^{3/2}(\hat{\delta}-\delta) \\ \sqrt{T}(\hat{\beta}-\beta) \end{bmatrix} = \begin{bmatrix} 1 & \frac{\sum_{t=1}^{T} t}{T^2} & \frac{\sum_{t=1}^{T} x_t}{T} \\ \frac{\sum_{t=1}^{T} t}{T^2} & \frac{\sum_{t=1}^{T} t^2}{T^3} & \frac{\sum_{t=1}^{T} t x_t}{T^2} \\ \frac{\sum_{t=1}^{T} x_t}{T} & \frac{\sum_{t=1}^{T} t x_t}{T^2} & \frac{\sum_{t=1}^{T} x_t^2}{T} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\sum_{t=1}^{T} \varepsilon_t}{\sqrt{T}} \\ \frac{\sum_{t=1}^{T} t\varepsilon_t}{T^{3/2}} \\ \frac{\sum_{t=1}^{T} x_t\varepsilon_t}{\sqrt{T}} \end{bmatrix} \tag{4.2.2}
$$

Para hallar la distribución asintótica se aplican los siguientes supuestos:

Suposición 1. *Sea x_t un proceso strong mixing con coeficiente α de tamaño $r/(r-1)$, $r > 1$ y $\mu_t \equiv E[X_t]$, tal que $\sup X_t \in \{X_t\}$ y $\sup E|X_t|^{r+\delta} \leq k < \infty$, $\delta > 1 \,\forall\, t$. Y que $\bar{\mu}_T = \frac{1}{T}\sum_{t=1}^{T} E[X_t]$ con $\lim\limits_{T\longrightarrow\infty} \bar{\mu}_T = \mu$.*

Suposición 2. *Sea ε_t un proceso strong mixing con coeficiente α de tamaño $r/(r-1)$, $r > 1$, tal que $E[\varepsilon_t] = 0 \quad \forall t$, $\sup \varepsilon_t \in \{\epsilon_t\}$ y $\sup E|\varepsilon_t|^{r+\delta} \leq k < \infty$, $\delta > 0 \,\forall t$*

y con $\sigma^2_\varepsilon = \lim_{T\longrightarrow\infty} E\left[\frac{1}{T}S^2_\varepsilon\right]$ $,\sigma^2_\varepsilon > 0$[2]. *Y que para* x_t *y* ε_t *se cumple:* $E\left[x_t\varepsilon_t\right] = 0 \quad \forall t$ *y* $\sigma^2_{x\varepsilon} = \lim_{T\longrightarrow\infty} E\left[\frac{1}{T}S^2_{x\varepsilon}\right]$ $,\sigma^2_{x\varepsilon} > 0$[3].

Dado los supuestos 1 y 2 se procede a determinar las propiedades asintóticas de cada elemento de la ecuación matricial (4.2.2).

El término de la primera fila y primera columa de la primera matriz de la ecuación (4.2.2) es 1, tomando límites se tiene:

$$\lim_{T\longrightarrow\infty} 1 = 1$$

Mediante aplicación de la proposición 2.2.7 se tiene que:

$$1 \xrightarrow{\quad P \quad} 1 \tag{4.2.3}$$

Mientras que en la primera matriz el término $\sum_{t=1}^{T} t/T^2$ de la ecuación (4.2.2) se puede expresar como:

$$\frac{\sum_{t=1}^{T} t}{T^2} = \frac{1}{T^2} \times \left[\frac{T(T+1)}{2}\right] = \frac{1}{2} + \frac{1}{(2T)}$$

Y tomando límites:

$$\lim_{T\longrightarrow\infty} \left[\frac{\sum_{t=1}^{T} t}{T^2}\right] = \lim_{T\longrightarrow\infty} \left[\frac{1}{2} + \frac{1}{(2T)}\right] = \frac{1}{2}$$

Por aplicación directa de la proposición 2.2.7:

$$\frac{\sum_{t=1}^{T} t}{T^2} \xrightarrow{\quad P \quad} \frac{1}{2} \tag{4.2.4}$$

De manera análoga $\sum_{t=1}^{T} t^2/T^3$ se puede expresar como:

$$\frac{\sum_{t=1}^{T} t^2}{T^3} = \frac{1}{T^3} \times \left[\frac{T(T+1)(2T+1)}{6}\right] = \frac{1}{3} + \frac{1}{(2T)} + \frac{1}{6T^2}$$

[2]Se define $S^2_\varepsilon = \sum_{t=1}^{T} var\varepsilon_t$ y $S^2_{x\varepsilon} = \sum_{t=1}^{T} varx_t\varepsilon_t$.

[3]Los términos σ^2_ε y $\sigma^2_{x\varepsilon}$ reciben el nombre de varianza a largo plazo.

Tomando límites:

$$\lim_{T\longrightarrow\infty}\left[\frac{\sum_{t=1}^{T}t^2}{T^3}\right]=\lim_{T\longrightarrow\infty}\left[\frac{1}{3}+\frac{1}{(2T)}+\frac{1}{6T^2}\right]=\frac{1}{3}$$

Aplicando la proposición 2.2.7 :

$$\frac{\sum_{t=1}^{T}t^2}{T^3}\xrightarrow{P}\frac{1}{3} \tag{4.2.5}$$

Mediante la suposición 1, x_t sigue un proceso α- mixing con coeficiente mixing $r/(r-1),\ r>1$ y satisface los requisitos del teorema 2.4.1[4]. De modo que para $\sum_{t=1}^{T}x_t/T$ en la ecuación (4.2.2) se tiene:

$$\begin{aligned}
&\frac{\sum_{t=1}^{T}x_t}{T}-\bar{\mu}_T\xrightarrow{a.s}0 && ,\ aplicación\ del\ teorema\ 2.4.1\\
&\frac{\sum_{t=1}^{T}x_t}{T}\xrightarrow{a.s}\mu && ,\ proposición\ 2.2.14\\
&\frac{\sum_{t=1}^{T}x_t}{T}\xrightarrow{p}\mu && ,\ aplicación\ de\ 2.2.5
\end{aligned}$$

de modo que:

$$\frac{\sum_{t=1}^{T}x_t}{T}\xrightarrow{p}\mu \tag{4.2.6}$$

De la proposición (2.4.1) se tiene que si x_t es $\alpha-$mixing con coeficiente mixing $r/(r-1),\ r>1$ entonces x_t^2 es $\alpha-$mixing con coeficiente mixing $r/(r-1),\ r>1$. Adicionalmente se define $\bar{\mu}_T^2=\frac{1}{T}\sum_{t=1}^{T}E\left[X_t^2\right]$ y dado que $\sup E\left|X_t\right|^{r+\delta}\leq k<\infty\ ,\delta>1\ \forall t$ permite la aplicación de la proposición (A.2.5) de modo que $\lim_{T\longrightarrow\infty}\bar{\mu}_T^2=\mu^2$. Entonces para $\sum_{t=1}^{T}x_t^2/T$ se tiene:

[4]Notar que $E\left|X_t\right|^{r+\delta}\leq k<\infty\ ,\delta>1\ \ r>1\,\forall t$ implica que$E\left|X_t\right|^{r+\delta}\leq k<\infty\ ,\delta>0\ \ r>1\,\forall t$ por proposición (2.2.2).

$$\frac{\sum_{t=1}^{T} x_t^2}{T} - \bar{\mu}_T^2 \xrightarrow{a.s} 0 \qquad , \textit{aplicación del teorema } 2.4.1$$
$$\frac{\sum_{t=1}^{T} x_t^2}{T} \xrightarrow{a.s} \mu^2 \qquad , \textit{proposición } 2.2.14$$
$$\frac{\sum_{t=1}^{T} x_t^2}{T} \xrightarrow{p} \mu^2 \qquad , \textit{aplicación de } 2.2.4$$

Por lo tanto:

$$\frac{\sum_{t=1}^{T} x_t}{T} \xrightarrow{p} \mu^2 \tag{4.2.7}$$

Asímismo, la expresión $\sum_{t=1}^{T} tx_t/T^2$ se puede expresar de la siguiente manera:

$$\begin{aligned}
\frac{\sum_{t=1}^{T} tx_t}{T^2} &= \frac{1}{T^2}\left[x_1 + 2x_2 + \ldots + Tx_T\right] \\
&= \frac{1}{T^2}\Big[x_1 + 2x_2 + \ldots + Tx_T + \big(x_2 + x_3 + \ldots + x_T\big) - \big(x_2 + x_3 + \ldots + x_T\big)\Big] \\
&= \frac{\sum_{t=1}^{T} x_t}{T^2} + \frac{1}{T^2}\Big[x_2 + 2x_3 \ldots + (T-1)\, x_T\Big] \\
&= \frac{\sum_{t=1}^{T} x_t}{T^2} + \frac{1}{T^2}\Big[x_2 + 2x_3 \ldots + (T-1)\, x_T + (x_3 + \ldots + x_T) - (x_3 + \ldots + x_T)\Big] \\
&= \frac{\sum_{t=1}^{T} x_t}{T^2} + \frac{\sum_{t=2}^{T} x_t}{T^2} + \frac{1}{T^2}\Big[x_3 + x_4 \ldots + (T-2)\, x_T\Big] \\
&= \vdots \\
&= \frac{1}{T^2} \times \sum_{t=1}^{T} x_t + \frac{1}{T^2} \times \sum_{t=2}^{T} x_t + \ldots + \frac{1}{T^2} \times \sum_{t=T-1}^{T} x_t + \frac{1}{T^2} \times x_t
\end{aligned}$$

Dado que por la suposición 1 x_t cumple con $\sup E\left|X_t\right|^{r+\delta} \leq k < \infty$, $\delta > 1$ $r > 1 \, \forall \, t$, lo cual implica que x_t es as acotada en L_p y por la proposición 2.2.5 se tiene que x_t es acotada estocásticamente de modo que $x_t = O_p(1) \;\; \forall \, t$. Por otro lado, $\lim_{T \longrightarrow \infty} \frac{1}{T^2} = 0$ y por la aplicación directa de la definición A.1.2 se tiene que $\frac{1}{T^2} = o(1)$. Entonces:

$$
\begin{aligned}
\frac{\sum_{t=1}^{T} t x_t}{T^2} =& \frac{1}{T^2} \times \sum_{t=1}^{T} x_t + \frac{1}{T^2} \times \sum_{t=2}^{T} x_t + \ldots + \frac{1}{T^2} \times \sum_{t=T-1}^{T} x_t + \frac{1}{T^2} \times x_t \\
=& o(1) \times O_p(1) + o(1) \times O_p(1) + \ldots + o(1) \times O_p(1) \\
=& o_p(1) \times O_p(1) + o_p(1) \times O_p(1) + \ldots + o_p(1) \times O_p(1), \; por\, A.1.8 \\
=& o_p(1) + o_p(1) + \ldots + o_p(1), \; por\, A.1.4 \\
=& o_p(1), \; por\, A.1.2
\end{aligned}
$$

Lo cual implica que:

$$
\frac{\sum_{t=1}^{T} t x_t}{T^2} \xrightarrow{p} 0 \tag{4.2.8}
$$

Aplicando la suposición 2 se tiene que la variable ε_t cumple con los supuestos del teorema (2.4.2) dado que $\sup E\left|\varepsilon_t\right|^{r+\delta} \leq k < \infty$ $, \delta > 1,\; r > 1 \;\forall\; t$ establece que la sucesión de expectativas de ε_t es acotada e integrable de orden $P = r + \delta$, lo que implica que $\lim\sup E\left|\varepsilon_t\right|^{r+\delta} < \infty$ $, \delta > 1,\; r > 1$. Mientras que $\sum_{i\in\mathbb{N}} \alpha(i)^{1-2/\beta} < \infty$ del teorema (2.4.2) se cumple si $\alpha(i) = O\left(m^{\frac{-r}{(r-2)}-\varepsilon}\right)$[5], por la definición (2.4.4) el tamaño de los coeficientes α-mixing ε_t en la suposición 1 es $\alpha(t) = O\left(m^{\frac{-r}{(r-1)}-\varepsilon}\right)$ y por la proposición (A.1.3) implica que $\alpha(t) = O\left(m^{\frac{-r}{(r-2)}-\varepsilon}\right)$ y se cumple el supuesto del teorema (2.4.2).

De modo que para el término $\sum_{t=1}^{T} \varepsilon_t/\sqrt{T}$ de la segunda matriz del lado derecho de la ecuación (4.2.2) se tiene que:

[5] Véase Phillips (1987) p-280.

$$\frac{\sum_{t=1}^{T}\varepsilon_t}{\sigma_\varepsilon\sqrt{T}} \xrightarrow{d} W(t) \qquad , \textit{teorema } 2.4.2$$
$$\frac{\sum_{t=1}^{T}\varepsilon_t}{\sigma_\varepsilon\sqrt{T}} \xrightarrow{d} N(0,1) \qquad , \textit{corolario } 2.4.1$$
$$\frac{\sum_{t=1}^{T}\varepsilon_t}{\sqrt{T}} \xrightarrow{d} \sigma_\varepsilon N(0,1) \qquad , \textit{proposición } 2.2.11$$

En consecuencia:

$$\frac{\sum_{t=1}^{T}\varepsilon_t}{\sqrt{T}} \xrightarrow{d} \sigma_\varepsilon N(0,1) \tag{4.2.9}$$

Para el término $\sum_{t=1}^{T} t\varepsilon_t/T^{3/2}$ de la segunda matriz del lado derecho de la ecuación (4.2.2), se define la variable $Z_t = Z_{t-1} + \varepsilon_t$ con $Z_0 = 0$, realizando reemplazos sucesivos se tiene que $Z_t = \sum_{t=1}^{T}\varepsilon_t$. De lo anterior se deriva la siguiente expresión:

$$\begin{aligned}
\frac{\sum_{t=1}^{T} Z_{t-1}}{T} &= \frac{1}{T}\Big[0 + \varepsilon_1 + (\varepsilon_1+\varepsilon_2) + \ldots + (\varepsilon_1+\varepsilon_2+\ldots+\varepsilon_{T-1})\Big] \\
&= \frac{1}{T}\Big[(T-1)\,\varepsilon_1 + (T-2)\,\varepsilon_2 + \ldots + \varepsilon_{T-1}\Big] \\
&= \frac{1}{T}\sum_{t=1}^{T}(T-t)\,\varepsilon_t \\
&= \sum_{t=1}^{T}\varepsilon_t + \frac{1}{T}\sum_{t=1}^{T} t\varepsilon_t
\end{aligned}$$

Desplazando términos se tiene que:

$$\frac{1}{T}\sum_{t=1}^{T} t\varepsilon_t = \sum_{t=1}^{T}\varepsilon_t - \frac{1}{T}\sum_{t=1}^{T} Z_{t-1}$$

Y Multiplicando ambos lados por $1/\sqrt{T}$ se obtiene:

$$\frac{1}{T^{3/2}}\sum_{t=1}^{T} t\varepsilon_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t - \frac{1}{T^{3/2}}\sum_{t=1}^{T} Z_{t-1}$$

Previamente bajo la suposición 2 se demosotró que $\sum_{t=1}^{T} \varepsilon_t/\sqrt{T} \xrightarrow{d} \sigma_\varepsilon W(1)$, por otro lado se tiene que $\sum_{t=1}^{T} Z_{t-1}/T^{3/2} \xrightarrow{d} \sigma_\varepsilon \int_0^1 W(s)ds$ [6]. Entonces:

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} t\varepsilon_t \xrightarrow{d} \sigma_\varepsilon W(1) - \sigma_\varepsilon \int_0^1 W(s)ds = \int_0^1 sdW(s)$$

Y por proposición (2.4.3) se tiene que:

$$\int_0^1 sdW(s) \sim N(0, 1/3)$$

De modo que:

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} t\varepsilon_t \xrightarrow{d} N(0, 1/3) \tag{4.2.10}$$

De la suposición 1 se tiene que x_t es $\alpha-$mixing con coeficientes mixing de tamaño $r/(r-1)$, $r > 1$ y de la proposición 2 ε_t es $\alpha-$mixing con coeficientes mixing de tamaño $r/(r-1)$, $r > 1$ y mediante proposición (2.4.2) se tiene que $x_t\varepsilon_t$ es $\alpha-$mixing con coeficientes mixing de tamaño $r/(r-1)$, $r > 1$. Y sea $X = \sup\left|X_t\right|^{r+\delta}$ y $Y = \sup\left|\varepsilon_t\right|^{r+\delta}$ respectivamente. Entonces se tiene que:

[6] Véase Hassler (2016) p-326 para una demostración de este resultado.

$$\begin{aligned}
E\left|XY\right| &= E\left[\sup\left|X_t\right|^{r+\delta}\sup\left|\varepsilon_t\right|^{r+\delta}\right] \\
&\geq E\left[\sup\left|X_t\varepsilon_t\right|^{r+\delta}\right] && \text{, } \mathit{por}\ \sup XY \leq \sup X \cdot \sup Y \\
&\geq \sup E\left|X_t\varepsilon_t\right|^{r+\delta} && \text{, } \mathit{ya\,que}\ \sup E\left[X\right] \leq E\left[\sup X\right] \\
E\left|XY\right| &\leq \sqrt{E\left|X\right|^2} \times \sqrt{E\left|Y\right|^2} && \text{, } \mathit{por}\ (2.\,2.\,1)\ \mathit{con}\, p = q = 2 \\
&= \left[-1\cdot\sqrt{E\left|X\right|^2}\right]\cdot\left[-1\cdot\sqrt{E\left|Y\right|^2}\right] \\
&\leq E\left[-\sqrt{\left|X\right|^2}\right] \times E\left[-\sqrt{\left|Y\right|^2}\right] && \text{, } \mathit{proposición}\, 2.\,2.\,3 \\
&\leq E\big[-\left|X\right|\big] \times E\big[-\left|Y\right|\big] \\
&\leq E\left[X\right] \times E\left[Y\right] && \text{, } \mathit{dado\,que}\ -\left|X\right| \leq X \\
&= E\Big[\sup\left|X_t\right|^{r+\delta}\Big] \times E\Big[\sup\left|\varepsilon_t\right|^{r+\delta}\Big] \\
E\left|XY\right| &< \infty && \text{, } \sup X_t \in \{X_t\}\ \text{, } \sup\varepsilon_t \in \{\epsilon_t\} \\
\sup E\left|X_t\varepsilon_t\right|^{r+\delta} &< \infty && (4.2.11)
\end{aligned}$$

Adicionalmente de la suposición 2 se cumple que $E\big[x_t\varepsilon_t\big] = 0\ \forall t$ y $\sigma_{x\varepsilon}^2 = \lim_{T\longrightarrow\infty} E\left[\frac{1}{T}S_{x\varepsilon}^2\right]$, $\sigma_{x\varepsilon}^2 > 0$ y junto a lo anterior se satisfacen los supuestos del teorema (2.4.2) para $x_t\varepsilon_t$. De modo que para $\sum_{t=1}^{T} x_t\varepsilon_t/\sqrt{T}$, se tiene:

$$\begin{aligned}
&\frac{\sum_{t=1}^{T} x_t\varepsilon_t}{\sigma_{x\epsilon}\sqrt{T}} \xrightarrow{d} W\left(t\right) && \text{, } \mathit{teorema}\ 2.\,4.\,2 \\
&\frac{\sum_{t=1}^{T} x_t\varepsilon_t}{\sigma_{x\epsilon}\sqrt{T}} \xrightarrow{d} N\left(0,1\right) && \text{, } \mathit{corolario}\, 2.\,4.\,1 \\
&\frac{\sum_{t=1}^{T} \varepsilon_t}{\sqrt{T}} \xrightarrow{d} \sigma_{x\epsilon}N(0,1) && \text{, } \mathit{proposición}\, 2.\,2.\,11
\end{aligned}$$

Por lo tanto:

$$\frac{\sum_{t=1}^{T} \varepsilon_t}{\sqrt{T}} \xrightarrow{d} \sigma_{x\epsilon}N(0,1) \qquad (4.2.12)$$

Aplicando los resultados de las ecuaciones (4.2.3), (4.2.4), (4.2.5), (4.2.6), (4.2.7), (4.2.8), (4.2.9), (4.2.10) y (4.2.12) en la ecuación matricial (4.2.2) se tiene que:

$$\begin{bmatrix} \sqrt{T}\left(\hat{\alpha}-\alpha\right) \\ T^{3/2}\left(\hat{\delta}-\delta\right) \\ \sqrt{T}\left(\hat{\beta}-\beta\right) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} 1 & \frac{1}{2} & \mu \\ \frac{1}{2} & \frac{1}{3} & 0 \\ \mu & 0 & \mu^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{\varepsilon} N(0,1) \\ N\left(0,1/3\right) \\ \sigma_{x\epsilon} N(0,1) \end{bmatrix} \tag{4.2.13}$$

Ahora se procede a hallar la matriz inversa[7] de la ecuación matricial (4.2.13). Primero se halla el determinante:

$$\begin{aligned} \begin{vmatrix} 1 & \frac{1}{2} & \mu \\ \frac{1}{2} & \frac{1}{3} & 0 \\ \mu & 0 & \mu^2 \end{vmatrix} &= 1 \times \begin{vmatrix} \frac{1}{3} & 0 \\ 0 & \mu^2 \end{vmatrix} - \frac{1}{2} \times \begin{vmatrix} \frac{1}{2} & 0 \\ \mu & \mu^2 \end{vmatrix} + \frac{1}{\mu} \times \begin{vmatrix} \frac{1}{2} & \frac{1}{3} \\ \mu & 0 \end{vmatrix} \\ &= \frac{\mu^2}{3} - \frac{\mu^2}{4} - \frac{[\mu]^2}{3} \\ &= \frac{\mu^2}{12} - \frac{[\mu]^2}{3} \\ &= \frac{\mu^2 - 4\,[\mu]^2}{12} \end{aligned}$$

Luego se hallan los cofactores de cada término de la primera matriz de la ecuación matricial (4.2.13):

[7]Véase el Apéndice A.3 para una revisión de los resultados básicos inversas de matrices.

$$A_{11} = (-1)^{1+1} \begin{vmatrix} \frac{1}{3} & 0 \\ 0 & \mu^2 \end{vmatrix} = 1 \times \left[\frac{1}{3} \times \mu^2 - 0\right] = \frac{\mu^2}{3}$$

$$A_{12} = (-1)^{1+2} \begin{vmatrix} \frac{1}{2} & 0 \\ \mu & \mu^2 \end{vmatrix} = -1 \times \left[\frac{1}{2} \times \mu^2 - 0 \times \mu\right] = -\frac{\mu^2}{2}$$

$$A_{13} = (-1)^{1+3} \begin{vmatrix} \frac{1}{2} & \frac{1}{3} \\ \mu & 0 \end{vmatrix} = 1 \times \left[\frac{1}{2} \times 0 - \frac{1}{3} \times \mu\right] = -\frac{\mu}{3}$$

$$A_{21} = (-1)^{2+1} \begin{vmatrix} \frac{1}{2} & \mu \\ 0 & \mu^2 \end{vmatrix} = -1 \times \left[\frac{1}{2} \times \mu^2 - 0 \times \mu\right] = -\frac{\mu^2}{2}$$

$$A_{22} = (-1)^{2+2} \begin{vmatrix} 1 & \mu \\ \mu & \mu^2 \end{vmatrix} = 1 \times \left[1 \times \mu^2 - \mu \times \mu\right] = \mu^2 - [\mu]^2$$

$$A_{23} = (-1)^{2+3} \begin{vmatrix} 1 & \frac{1}{2} \\ \mu & 0 \end{vmatrix} = -1 \times \left[1 \times 0 - \mu \times \frac{1}{2}\right] = \frac{\mu}{2}$$

$$A_{31} = (-1)^{3+1} \begin{vmatrix} \frac{1}{2} & \mu \\ \frac{1}{3} & 0 \end{vmatrix} = 1 \times \left[\frac{1}{2} \times 0 - \frac{1}{3} \times \mu\right] = -\frac{\mu}{3}$$

$$A_{32} = (-1)^{3+2} \begin{vmatrix} 1 & \mu \\ \frac{1}{2} & 0 \end{vmatrix} = -1 \times \left[1 \times 0 - \mu \times \frac{1}{2}\right] = \frac{\mu}{2}$$

$$A_{33} = (-1)^{3+3} \begin{vmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{vmatrix} = 1 \times \left[1 \times \frac{1}{3} - \frac{1}{2} \times \frac{1}{2}\right] = -\frac{2}{3}$$

De modo que la matriz de cofactores es:

$$[\mathbf{A}_{ij}] = \begin{bmatrix} \frac{\mu^2}{3} & -\frac{\mu^2}{2} & -\frac{\mu}{3} \\ -\frac{\mu^2}{2} & \mu^2 - [\mu]^2 & \frac{\mu}{2} \\ -\frac{\mu}{3} & \frac{\mu}{2} & -\frac{2}{3} \end{bmatrix}$$

Y la matriz adjunta:

$$adj\mathbf{A} = \begin{bmatrix} \frac{\mu^2}{3} & -\frac{\mu^2}{2} & -\frac{\mu}{3} \\ -\frac{\mu^2}{2} & \mu^2 - [\mu]^2 & \frac{\mu}{2} \\ -\frac{\mu}{3} & \frac{\mu}{2} & -\frac{2}{3} \end{bmatrix}$$

Y la inversa es según la ecuación:

$$\begin{bmatrix} 1 & \frac{1}{2} & \mu \\ \frac{1}{2} & \frac{1}{3} & 0 \\ \mu & 0 & \mu^2 \end{bmatrix}^{-1} = \frac{1}{\frac{\mu^2 - 4\,[\mu]^2}{12}} \times \begin{bmatrix} \frac{\mu^2}{3} & -\frac{\mu^2}{2} & -\frac{\mu}{3} \\ -\frac{\mu^2}{2} & \mu^2 - [\mu]^2 & \frac{\mu}{2} \\ -\frac{\mu}{3} & \frac{\mu}{2} & -\frac{2}{3} \end{bmatrix}$$

$$= \frac{12}{\triangle} \times \begin{bmatrix} \frac{\mu^2}{3} & -\frac{\mu^2}{2} & -\frac{\mu}{3} \\ -\frac{\mu^2}{2} & \mu^2 - [\mu]^2 & \frac{\mu}{2} \\ -\frac{\mu}{3} & \frac{\mu}{2} & -\frac{2}{3} \end{bmatrix}, \triangle = \mu^2 - 4\,[\mu]^2$$

$$= \begin{bmatrix} \frac{4\mu^2}{\triangle} & -\frac{6\mu^2}{\triangle} & -\frac{4\mu}{\triangle} \\ -\frac{6\mu^2}{\triangle} & \frac{12\left[\mu^2 - [\mu]^2\right]}{\triangle} & \frac{6\mu}{\triangle} \\ -\frac{4\mu}{\triangle} & \frac{6\mu}{\triangle} & -\frac{8}{\triangle} \end{bmatrix}$$

En consencuencia la ecuación matricial (4.2.13) queda como:

$$\begin{bmatrix} \sqrt{T}\left(\hat{\alpha}-\alpha\right) \\ T^{3/2}\left(\hat{\delta}-\delta\right) \\ \sqrt{T}\left(\hat{\beta}-\beta\right) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \dfrac{4\mu^2}{\triangle} & -\dfrac{6\mu^2}{\triangle} & -\dfrac{4\mu}{\triangle} \\ -\dfrac{6\mu^2}{\triangle} & \dfrac{12\left[\mu^2-[\mu]^2\right]}{\triangle} & \dfrac{6\mu}{\triangle} \\ -\dfrac{4\mu}{\triangle} & \dfrac{6\mu}{\triangle} & -\dfrac{8}{\triangle} \end{bmatrix} \begin{bmatrix} \sigma_\varepsilon N(0,1) \\ N\left(0,1/3\right) \\ \sigma_{x\epsilon} N(0,1) \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{4\mu^2}{\triangle} & -\dfrac{6\mu}{\triangle} & -\dfrac{4\mu}{\triangle} \\ -\dfrac{6\mu^2}{\triangle} & \dfrac{12\left[\mu^2-[\mu]^2\right]}{\triangle} & \dfrac{6\mu}{\triangle} \\ -\dfrac{4\mu}{\triangle} & \dfrac{6\mu}{\triangle} & -\dfrac{8}{\triangle} \end{bmatrix} \begin{bmatrix} \sigma_\varepsilon & 0 & 0 \\ 0 & \dfrac{1}{3} & 0 \\ 0 & 0 & \sigma_{x\epsilon} \end{bmatrix} \begin{bmatrix} N(0,1) \\ N\left(0,1\right) \\ N(0,1) \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{4\mu^2\sigma_\varepsilon}{\triangle} & -\dfrac{2\mu^2}{\triangle} & -\dfrac{4\mu\sigma_{x\epsilon}}{\triangle} \\ -\dfrac{6\mu^2\sigma_\varepsilon}{\triangle} & \dfrac{4\left[\mu^2-[\mu]^2\right]}{\triangle} & \dfrac{6\mu\sigma_{x\epsilon}}{\triangle} \\ -\dfrac{4\mu\sigma_\varepsilon}{\triangle} & \dfrac{2\mu}{\triangle} & -\dfrac{8\sigma_{x\epsilon}}{\triangle} \end{bmatrix} \begin{bmatrix} N(0,1) \\ N\left(0,1\right) \\ N(0,1) \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{4\mu^2\sigma_\varepsilon}{\triangle} N(0,1) - \dfrac{2\mu^2}{\triangle} N\left(0,1\right) - \dfrac{4\mu\sigma_{x\epsilon}}{\triangle} N(0,1) \\ -\dfrac{6\mu^2\sigma_\varepsilon}{\triangle} N\left(0,1\right) + \dfrac{4\left[\mu^2-[\mu]^2\right]}{\triangle} N\left(0,1\right) + \dfrac{6\mu\sigma_{x\epsilon}}{\triangle} N\left(0,1\right) \\ -\dfrac{4\mu\sigma_\varepsilon}{\triangle} N(0,1) + \dfrac{2\mu}{\triangle} N\left(0,1\right) - \dfrac{8\sigma_{x\epsilon}}{\triangle} N(0,1) \end{bmatrix}$$

Y por aplicación de la proposición (A.4.1) se tiene:

$$\begin{bmatrix} \sqrt{T}\left(\hat{\alpha}-\alpha\right) \\ T^{3/2}\left(\hat{\delta}-\delta\right) \\ \sqrt{T}\left(\hat{\beta}-\beta\right) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} N\left(0,\left[\dfrac{4\mu^2\sigma_\varepsilon}{\triangle}\right]^2\right) + N\left(0,\left[-\dfrac{2\mu^2}{\triangle}\right]^2\right) + N\left(0,\left[-\dfrac{4\mu\sigma_{x\epsilon}}{\triangle}\right]^2\right) \\ N\left(0,\left[-\dfrac{6\mu^2\sigma_\varepsilon}{\triangle}\right]^2\right) + N\left(0,\left[\dfrac{4\left[\mu^2-[\mu]^2\right]}{\triangle}\right]^2\right) + N\left(0,\left[\dfrac{6\mu\sigma_{x\epsilon}}{\triangle}\right]^2\right) \\ N\left(0,\left[-\dfrac{4\mu\sigma_\varepsilon}{\triangle}\right]^2\right) + N\left(0,\left[\dfrac{2\mu}{\triangle}\right]^2\right) + N\left(0,\left[-\dfrac{8\sigma_{x\epsilon}}{\triangle}\right]^2\right) \end{bmatrix}$$

Aplicando la proposición (A.4.2), de modo que:

$$\begin{bmatrix} \sqrt{T}\big(\hat{\alpha}-\alpha\big) \\ T^{3/2}\big(\hat{\delta}-\delta\big) \\ \sqrt{T}\big(\hat{\beta}-\beta\big) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} N\left(0, \left[\frac{4\mu^2\sigma_\varepsilon}{\triangle}\right]^2 + \left[-\frac{2\mu^2}{\triangle}\right]^2 + \left[-\frac{4\mu\sigma_{x\epsilon}}{\triangle}\right]^2\right) \\ N\left(0, \left[-\frac{6\mu^2\sigma_\varepsilon}{\triangle}\right]^2 + \left[\frac{4\left[\mu^2-[\mu]^2\right]}{\triangle}\right]^2 + \left[\frac{6\mu\sigma_{x\epsilon}}{\triangle}\right]^2\right) \\ N\left(0, \left[-\frac{4\mu\sigma_\varepsilon}{\triangle}\right]^2 + \left[\frac{2\mu}{\triangle}\right]^2 + \left[-\frac{8\sigma_{x\epsilon}}{\triangle}\right]^2\right) \end{bmatrix} \tag{4.2.14}$$

De la ecuaciòn matricial (4.2.14) se evidencian tres hechos importantes, primero que cada uno de los elementos los estimadores de $(\mathbf{b}-\beta)$ para los estimadores $\hat{\alpha}$, $\hat{\delta}$, $\hat{\beta}$ del modelo de regresión (4.1.1) convergen a una distribución normal con media 0 pero con diferentes varianzas, segundo cada una de las varianzas de $(\mathbf{b}-\beta)$ en la ecuación matricial (4.2.14) depende no solo de los propios parámetros poblacionales sino de los parámetros de otras variables, y tercero cada una de las variables en $(\mathbf{b}-\beta)$ converge a la distribución normal a una tasa distinta para el caso de los estimadores de $\hat{\alpha}$ y $\hat{\beta}$ convergencen a una tasa $\sqrt{T}$ mientras que $\hat{\delta}$ converge a una tasa $T^{3/2}$.

Capítulo 5

Conclusiones y recomendaciones

5.1. Conclusión.

La determinación de la distribución de los estimadores en modelos de regresión juega un papel preponderante en los procedimientos inferenciales en los modelos de regresión como se evidenció en la sección 3.4 del capítulo 3 del presente trabajo.

Los tres puntos del último párrafo de la sección 4.2 tienen consecuencias importantes en el proceso de inferencia en los modelos de regresión lineal cuandos las variables siguen procesos strong mixing con componentes tendenciales dado que en el proceso inferencial tradicional de regresiones con variables explicativas determinísticas, el vector $(\mathbf{b}-\beta)$ sigue distribución normal $N(0,1)$ de modo que es posible usar tablas estadísticas convencionales para los diversos test de significancia individual y conjunta, mientras que los estimadores de los regresores strong mixing se distribuen de acuerdo a (4.2.14). De modo que aplicar estos procedimientos en regresiones con variables strong mixing puede llevar erroneamente a establecer relaciones estadísticas cuando puede no existir ninguna. Por otro lado, dado que las varianzas de las distribuciones para las variables $\hat{\alpha}$, $\hat{\delta}$, $\hat{\beta}$ en la ecuación matricial (4.2.14) dependen de parametros poblacionales, es necesario estimar estos parámetros a la hora de realizar el proceso inferencial sobre los estimadores $\hat{\alpha}$, $\hat{\delta}$, $\hat{\beta}$.

De manera general los procesos de inferencia estadística en los modelos de regresión con datos observacionales, esto datos provenientes de la observación pasiva en contraste con datos que provienen de muestras o experimentos, no deben de ser llevados de forma mecánica, y de manera particular con regresores que siguen procesos dependientes strong mixing y componentes tendenciales dado que se corre el riesgo de cometer errores tipo I o tipo II y asignar una relación estadística cuando en realidad no existe ninguna.

5.2. Recomendación

Se recomienda la aplicación de un análisis minucioso a la hora de aplicar modelos de regresiones sobre datos de series temporales, dado que sobre las propiedades que pueda tener cada serie en el modelo de regresión va a depender en gran manera la distribución de los estimadores que son esenciables para establecer las distribuciones de las pruebas estadísticas posteriores, que se realizan en el proceso de inferencia, ignorar este punto puede llevar realizar procedimientos inferenciales inválido. Por otro lado se recomienda usar la distribución de los estimadores obtenidos en el presente trabajo dado que los procesos strong-mixing son bastante generales y permiten tanto su aplicación para series estacionarias y no estacionarias que son de uso común en el trabajo aplicado.

Apéndice A

Fórmulas y procedimientos

A.1. Notación O-grande, o-pequeña, O_p y o_p

Definición A.1.1. Notación $O-$grande.[1] Una función $f(x)$ es de orden $g(x)$ si[2]:

$$\lim_{n \longrightarrow \infty} \frac{f(x)}{g(x)} = x_o$$

Para un $x_0 \in \mathbb{R}$ y se representa como: $f(x) = O(g(x))$.

Definición A.1.2. Notación $o-$pequeña[3]. Una función $f(x)$ es de orden cero relativo a $g(x)$ si:

$$\lim_{n \longrightarrow \infty} \frac{f(x)}{g(x)} = 0$$

Y se representa como: $f(x) = o(g(x))$.

Definición A.1.3. Notación O_p-grande. Se dice que $\{X_n\}$ es de orden en probabilidad $O(f(N))$ si para cada $\epsilon > 0$ existe un $A_\epsilon > 0$ tal que:

$$P\left[\frac{X_n}{f(N)} \leq A_\epsilon\right] \geq 1 - \epsilon$$

[1] La notación $O-$ grande y $o-$pequeña se aplica indistintamente tanto a funciones como a sucesiones, recordando que una sucesión es una función cuyo dominio es $\mathbb{N}$

[2] Véae Shynk (2013) p-640.

[3] Ibid, p-641.

Y se denota como $X_n = O(f(N))$.

Definición A.1.4. Notación o_p-pequeña. Se dice que $\{X_n\}$ es de orden en probabilidad $o(f(N))$ si:

$$p\lim \frac{X_n}{f(N)} = 0$$

Y se denota como: $X_n = o(f(N))$.

Proposición A.1.1. *Sea $f = O(a_n)$, $g = O(b_n)$, $h = o(c_n)$ y $j = o(d_n)$, entonces se cumple que*[4]*:*

$$f + g = O\big(\text{máx}\{a_n, b_n\}\big) \tag{A.1.1}$$

$$h + j = o\big(\text{máx}\{c_n, d_n\}\big) \tag{A.1.2}$$

$$f + h = o(c_n) \tag{A.1.3}$$

$$f \times h = o(a_n c_n) \tag{A.1.4}$$

$$h \times j = o(c_n d_n) \tag{A.1.5}$$

$$f \times g = O(a_n b_n) \tag{A.1.6}$$

Estas mismas propiedades se aplican para O_p y o_p.

Proposición A.1.2. *Si $x_n = O(c_n)$ y $y_n = o(b_n)$. Entonces*[5]*:*

$$x_n = O_p(c_n) \tag{A.1.7}$$

$$y_n = o_p(b_n) \tag{A.1.8}$$

Proposición A.1.3. *Si $x_n = O(n^{\lambda})$ entonces para un $\delta > 0$, $x_n = O(n^{\lambda+\delta})$*[6]*.*

A.2. Resultados de Sucesiones y series

Proposición A.2.1. *El $\sup a_n$ existe si y solo si a_n es acotada desde arriba.*

[4] Véase Spanos 1986, pag-195.
[5] Véase Mittelhammer (2013) p-247.
[6] Véase White (2000) p-17.

Proposición A.2.2. *Si* $\sup a_n < \infty$ *entonces* $\limsup a_n < \infty$.

Proposición A.2.3. *Teorema de Bolzano-Weierstrass. Una sucesión acotada* $\{a_n\}$ *tiene una sub-sucesión* a_{n_k} *que es convergente.*

Proposición A.2.4. *Una sucesión*$\{a_n\}$ *es convergente si y solo si es acotada y tiene una sub-sucesión* a_{n_k} *que converge a un límite.*

Proposición A.2.5. *Sea* $\{a_n\}$ *una sucesión, y se define la siguiente sucesión:*

$$b_n = \frac{a_1 + a_2 + a_3 + \ldots + a_n}{n}$$

Si $a_n \longrightarrow k$ *entonces* $b_n \longrightarrow k$[7].

A.3. Resultados de álgebra lineal

Definición A.3.1. Matriz[8]. Una matriz es un arreglo bidimensional de elementos, que pueden ser números, funciones, sucesiones e incluso otras matrices. Una matriz $\mathbf{A}$ se denota como:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}$$

La matriz $\mathbf{A}$ tiene n filas y m columnas, de modo que se denota como $\mathbf{A}_{m \times n}$.

Definición A.3.2. determinante de una una matriz. El determinante de una matriz $\mathbf{A}$ es un escalar asociado a esta matriz y se denota como $|\mathbf{A}|$, para el caso de una matriz cuadrada de 2x2 se tiene:

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

[7]Véase Yau (2013) p-59.

[8]Esta sección se basa principalmente en Poole (2006), capítulo 3.

Para el caso de una matriz cuadrada 3x3:

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

$$= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Y de manera general:

$$|\mathbf{A}| = \sum_{j=1}^{n} (-1)^{1+j} |\mathbf{A}_{1j}|$$

Donde $|\mathbf{A}_{1j}|$ se el determinante de la sub-matriz que resulta de eliminar la primera fila y j-ésima columna, y recibe el nombre de menor. y la expresión $C_{1j} = (-1)^{1+j} |\mathbf{A}_{1j}|$recibe el nombre de cofactor.

Definición A.3.3. Matriz de confactores. Sea $\mathbf{A}$ una matriz de orden $m \times n$ y sea C_{1j} el cofactor de cada elemento a_{ij}del determinante de la matriz $\mathbf{A}$. Entonces la matriz que contiene todos los cofactores de la matriz $\mathbf{A}$ recibe el nombre de matriz de cofacter y se denota como $[\mathbf{A}_{ij}]$. Para el caso de una matriz 3x3 se tiene:

$$[\mathbf{A}_{ij}] = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}$$

Definición A.3.4. Adjunta de una matriz. La transpuesta de la matriz de cofactores $[\mathbf{A}_{ij}]$se demonina adjunta y se denota como $adj\mathbf{A}$.

Definición A.3.5. Inversa de una matriz. La inversa de una matriz se define como:

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}.\, adj\mathrm{A}$$

A.4. Distribución Normal

Definición A.4.1. Distribución Normal.[9] Una variable aleatoria X sigue una distribució normal con expectativa μ y varianza σ^2 si su función de densidad esta dada por:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-u)^2}{2\sigma^2}} \quad , \ -\infty < x < \infty$$

Y su función de distribución es:

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-(y-u)^2}{2\sigma^2}} dy \quad , \ -\infty < x < \infty$$

Y se denota como $X \sim N(\mu, \sigma^2)$.

Y una distribución normal con $\mu = 0$ y $\sigma^2 = 1$ es llamada distribución normal estándar si función de densidad y distribución toma la siguientes formas respectivamente:

$$\phi(x) = f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2}} \quad , \ -\infty < x < \infty$$

$$\Phi(x) = F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt \quad , \ -\infty < x < \infty$$

Y es denotada como $X \sim N(0,1)$.

Proposición A.4.1. *Si $X \sim N(\mu, \sigma^2)$ entonces para una constante $c \in \mathbb{R}$ se tiene que $aX \sim N(c\mu, c\sigma^2)$[10].*

Proposición A.4.2. *Sea $X \sim N(\mu_x, \sigma_x^2)$ y $Y \sim N(\mu_y, \sigma_y^2)$ de modo que X es independiente de Y, entonces se cumple que $X + Y \sim N(\mu_{x+y}, \sigma_{x+y}^2)$[11].*

[9] Una completa explicación de la distribución normal y sus propiedades se puede encontrar Ahsanullah (2014), p-8.

[10] Para una demostración de este resultado véase Khuri (2010), p-66.

[11] Una demostración de esta proposición se encuentra en Roussas (2003), p-161

Bibliografía

Ahsanullah, M., B. M. G. Kibria, and M. Shakil (2014). *Normal and student's t distributions and their applications.* Atlantis Studies in Probability and Statistics: volume 4. Paris : Atlantis Press.

Athreya, K. B. and S. N. Lahiri (2006). *Measure theory and probability theory.* Springer texts in statistics. New York : Springer.

Basu, A. K. (2012). *Measure theory and probability* (2 ed.). Prentice-Hall of India.

Bhattacharya, R. N. and E. C. Waymire (2007). *A basic course in probability theory.* Universitext. New York : Springer.

Bierens, H. J. (1994). *Topics in advanced econometrics : estimation, testing, and specification of cross-section and time series models.* Cambridge [England] ; New York, NY, USA : Cambridge University Press.

Bierens, H. J. (2004). *Introduction to the mathematical and statistical foundations of econometrics.* Themes in modern econometrics. Cambridge, UK ; New York : Cambridge University Press.

Billingsley, P. (1968). *Convergence of probability measures.* Wiley series in probability and mathematical statistics. New York, Wiley.

Billingsley, P. (1995). *Probability and measure.* (3 ed.). Wiley series in probability and mathematical statistics. New York : J. Wiley & Sons.

Borodin, A. N. and P. Salminen (2002). *Handbook of Brownian motion : facts and formulae.* (2 ed.). Probability and its applications. Basel ; Boston : Birkhäuser.

Borovkov, A. A., K. A. Borovkov, O. B. Borovkova, and P. S. Ruzankin (2013). *Probability theory.* Universitext. London : Springer.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probab. Surveys 2*, 107–144.

Bruckner, A. M., J. B. Bruckner, and B. S. Thomson (1997). *Real analysis.* Upper Saddle River, N.J. : Prentice-Hall.

Capasso, V. and D. Bakstein (2012). *An introduction to continuous-time stochastic processes : theory, models, and applications to finance, biology, and medicine.* (2 ed.). Modeling and simulation in science, engineering and technology. Boston : Birkhäuser.

Cinlar, E. (2011). *Probability and Stochastics.* Graduate texts in mathematics: 261. New York: Londn Springer.

Dudley, R. M. (2002). *Real analysis and probability.* Cambridge studies in advanced mathematics: 74. Cambridge, U.K. ; New York : Cambridge University Press.

Durlauf, S. and P. Phillips (1988). Trends versus random walks in time series analysis. *Econometrica 56*(6), 1333–54.

Entorf, H. (1997). Random walks with drifts: Nonsense regression and spurious fixed-effect estimation. *Journal of Econometrics 80*(2), 287 – 296.

Fischer, Hans, D. (2011). *A history of the central limit theorem : from classical to modern probability theory.* Sources and studies in the history of mathematics and physical sciences. New York ; London : Springer.

Golberg, M. and H. A. Cho (2003). *Introduction to Regression Analysis.* Computational Mechanics.

Greene, W. H. (2012). *Econometric analysis.* (2 ed.). Upper Saddle River, N.J. : Pearson/Prentice Hall.

Gut, A. (2013). *Probability : a graduate course.* Springer texts in statistics. New York : Springer Science+Business Media.

Halbert White, I. D. (1984). Nonlinear regression with dependent observations. *Econometrica 52*(1), 143–161.

Haldrup, N. (1994). The asymptotics of single-equation cointegration regressions with i(1) and i(2) variables. *Journal of Econometrics 63*(1), 153–181.

Hasseler, U. (2000). Simple regressions with linear time trends. *Journal of Time Series Analysis 21*(1), 27–32.

Hassler, U. (1996). Spurious regressions when stationary regressors are included. *Economics Letters 50*(1), 25 – 31.

Hassler, U. (2016). *Stochastic Processes and Calculus An Elementary Introduction with Applications.* Springer International Publishing.

Hayashi, F. (2000). *Econometrics.* Princeton, N.J. : Princeton University Press.

Herrndorf, N. (1984, 02). A functional central limit theorem for weakly dependent sequences of random variables. *Annals of probability 12*(1), 141–153.

Jiang, J. (2010). *Large sample techniques for statistics.* Springer texts in statistics. New York : Springer.

Joon Y. Park, P. C. B. P. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory 4*(3), 468–497.

Joon Y. Park, P. C. B. P. (1989). Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory 5*(1), 95–131.

Karlin, S. and H. M. Taylor (1975). *A first course in stochastic processes.* (2 ed.). New York : Academic Press.

Khoshnevisan, D. (2007). *Probability.* Graduate studies in mathematics: v. 80. Providence, R.I. : American Mathematical Society.

Khuri, A. I. (2010). *Linear model methodology.* Boca Raton : CRC Press.

Kobayashi, H., B. L. Mark, and W. Turin (2012). *Probability, random processes, and statistical analysis.* Cambridge ; New York : Cambridge University Press.

Krämer, W. and F. Marmol (2002). Ols-based asymptotic inference in linear regression models with trending regressors and ar(p)-disturbances. *Communications in Statistics - Theory and Methods 31*(2), 261–270.

Liero, H. and S. Zwanzig (2012). *Introduction to the theory of statistical inference.* Texts in statistical science. Boca Raton, FL : CRC Press.

Lindsey, J. K. (1996). *Parametric statistical inference.* Oxford : Clarendon Press.

Loeve, M. (1978). *Probability Theory II* (4 ed.). Springer-Verlag New York.

McLeish, D. L. (1975, 10). A maximal inequality and dependent strong laws. *Ann. Probab. 3*(5), 829–839.

Mishura, Y. (2008). *Stochastic calculus for fractional Brownian motion and related processes.* Lecture notes in mathematics: 1929. Berlin ; New York : Springer-Verlag.

Mittelhammer, R. (2013). *Mathematical statistics for economics and business.* (2 ed.). New York : Springer.

Noriega, A. E. and D. Ventosa-Santaulària (2007). Spurious regression and trending variables*. *Oxford Bulletin of Economics and Statistics 69*(3), 439–444.

Panik, M. J. (2005). *Advanced statistics from an elementary point of view.* Amsterdam ; Boston : Elsevier/Academic Press.

Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica 55*(2), 277–301.

Phillips, P. C. B. and S. N. Durlauf (1986). Multiple time series regression with integrated processes. *Review of Economic Studies 53*(4), 473–495.

Polansky, A. M. (2011). *Introduction to statistical limit theory.* Chapman & Hall/CRC texts in statistical science series. Boca Raton, FL : Chapman & Hall/CRC Press.

Poole, D. (2006.). *Student Solutions Manual for Poole's Linear Algebra : A Modern Introduction* (2nd ed. ed.). Brooks Cole. Paperback.

Roussas, G. G. (2013). *An introduction to probability and statistical inference.* London, U.K. ; San Diego, CA : Academic Press,.

Roussas, G. G. (2014). *An introduction to measure-theoretic probability.* (2 ed.). Kidlington, Oxford ; Waltham, MA : Academic Press.

Schilling, R. L. (2005). *Measures, integrals and martingales.* Cambridge : Cambridge University Press.

Seber, G. A. F. and A. J. Lee (2003). *Linear regression analysis.* (2 ed.). Wiley series in probability and statistics. Hoboken, NJ : John Wiley.

Sen, A. and M. S. Srivastava (1990). *Regression analysis : theory, methods, and applications.* Springer texts in statistics. New York, Berlin, Paris: Springer-Verlang.

Shao, J. (2003). *Mathematical statistics.* Springer texts in statistics. New York : Springer.

Shirali, S. and H. L. Vasudeva (2006). *Metric spaces.* London : Springer.

Shiryaev, A. N. (1996). *Probability.* Graduate Texts in Mathematics. Springer New York.

Shynk, J. J. (2013). *Probability, random variables, and random processes : theory and signal processing applications.* Hoboken, N.J. : Wiley.

Spanos, A. (1986). *Statistical Foundations of Econometric Modelling.* Cambridge University Press.

Spanos, A. (1999). *Probability theory and statistical inference : econometric modeling with observational data.* Cambridge, U.K. ; New York : Cambridge University Press.

Stewart, C. (2006). Spurious correlation of i(0) regressors in models with an i(1) dependent variable. *Economics Letters 91*(2), 184 – 189.

Stock, J., C. Sims, and M. Watson (1990). Inference in linear time series models with some unit roots. *Econometrica 58*(1), 113–144.

Westfall, P. H. and K. S. S. Henning (2013). *Understanding advanced statistical methods.* Texts in statistical science series. Boca Raton : CRC Press, Taylor & Francis roup.

White, H. (2000). *Asymptotic theory for econometricians.* Economic theory, econometrics, and mathematical economics. San Diego, Calif. ; London : Academic.

Yau, D. (2013). *A first course in analysis.* Singapore: World Scientific.

Yeh, J. and J. Yeh (2006). *Real analysis : theory of measure and integration.* (2 ed.). Hackensack, NJ : World Scientific.

Zhengyan, L. and L. Chuanrong (1996). *Limit Theory for Mixing Dependent Random Variables.* Springer Netherlands.